

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 12, 2005, 12:16:52 ; Search time 243 Seconds  
(without alignments)  
1876.966 Million cell updates/sec

Title: US-10-627-075-1

Perfect score: 1220

Sequence: 1 gagctctccagctaccgacgc.....ccaaaaagagagagagatc 1220

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 3392430 seqs, 186927314 residues

Total number of hits satisfying chosen parameters: 6784860

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications NA.New.\*

1: /cgn2\_6/ptodata/1/pubpna/US09\_NEW\_PUB.seq.\*  
2: /cgn2\_6/ptodata/1/pubpna/US06\_NEW\_PUB.seq.\*  
3: /cgn2\_6/ptodata/1/pubpna/US07\_NEW\_PUB.seq.\*  
4: /cgn2\_6/ptodata/1/pubpna/US08\_NEW\_PUB.seq.\*  
5: /cgn2\_6/ptodata/1/pubpna/PCT\_NEW\_PUB.seq.\*  
6: /cgn2\_6/ptodata/1/pubpna/US10\_NEW\_PUB.seq.\*  
7: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*  
8: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*  
9: /cgn2\_6/ptodata/1/pubpna/US11\_NEW\_PUB.seq.\*  
10: /cgn2\_6/ptodata/1/pubpna/US60\_NEW\_PUB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	44.6	3.7	96583	US-11-117-187-203	Sequence 203, App
2	42.4	3.5	1711	US-10-750-185-25798	Sequence 25798, A
3	41.6	3.4	1625	US-10-750-185-55754	Sequence 55754, A
4	40	3.3	1593	US-10-750-185-54114	Sequence 54114, A
5	39.4	3.2	496	US-10-131-826A-533	Sequence 533, App
6	39.2	3.2	156250	US-11-121-086-86	Sequence 86, App1
7	38.8	3.2	101786	US-11-117-187-199	Sequence 199, App1
8	38.6	3.2	159146	US-11-121-086-49	Sequence 49, App1
9	38.6	3.2	611587	US-11-117-187-209	Sequence 209, App
10	37.8	3.1	788	US-10-750-185-54249	Sequence 49249, A
11	37.8	3.1	989	US-10-750-185-49243	Sequence 49243, A
12	37.8	3.1	3631	US-10-750-185-61642	Sequence 61642, A
13	37.8	3.1	4071	US-10-793-626-3381	Sequence 3381, App
14	37.8	3.1	1082144	US-11-117-187-211	Sequence 211, App
15	37.4	3.1	1450	US-11-117-187-211	Sequence 211, App
16	37.4	3.1	246960	US-10-750-185-52397	Sequence 52397, A
17	37.4	3.1	2479	US-11-121-086-8	Sequence 8, App1
18	37	3.0	2479	US-10-955-054A-50	Sequence 50, App1
19	36.8	3.0	1289	US-10-750-185-40390	Sequence 40390, A
20	36.8	3.0	2420	US-11-186-284-102	Sequence 102, App
21	36.2	3.0	1595	US-11-186-284-102	Sequence 102, App
22	36.2	3.0	1967	US-10-750-185-47422	Sequence 47422, A
23	36.2	3.0	3926	US-10-750-185-27185	Sequence 27185, A
				US-10-793-626-4300	Sequence 4300, App

24	36.2	3.0	4189	US-10-793-626-3898	Sequence 3898, App
25	36.2	3.0	4339	US-10-909-125-801	Sequence 801, App
26	36	3.0	2356	US-10-821-234-815	Sequence 815, App
27	36	3.0	161874	US-11-121-086-75	Sequence 75, App1
28	35.8	2.9	1073	US-11-102-240-95	Sequence 95, App1
29	35.8	2.9	184000	US-11-121-086-37	Sequence 37, App1
30	35.8	2.9	189252	US-11-121-086-54	Sequence 54, App1
31	35.8	2.9	189993	US-11-121-086-78	Sequence 78, App1
32	35.6	2.9	794	US-10-750-185-62887	Sequence 62887, A
33	35.6	2.9	1619	US-10-750-185-63154	Sequence 63154, A
34	35.6	2.9	3657	US-10-750-185-43919	Sequence 43919, A
35	35.6	2.9	6450	US-11-091-668-3	Sequence 3, App1
36	35.4	2.9	958	US-11-179-411-5	Sequence 5, App1
37	35.4	2.9	1608	US-10-750-185-24686	Sequence 24686, A
38	35.4	2.9	164527	US-11-121-086-71	Sequence 71, App1
39	35.2	2.9	742	US-10-750-185-62008	Sequence 62008, A
40	35.2	2.9	1720	US-10-750-185-43422	Sequence 43422, A
41	35.2	2.9	21777	US-11-027-964-2	Sequence 2, App1
42	35.2	2.9	43948	US-10-949-720-393	Sequence 393, App
43	35	2.9	959	US-10-750-185-60341	Sequence 60341, A
44	35	2.9	1267	US-10-750-185-54507	Sequence 54507, A
45	35	2.9	1468	US-10-750-185-57297	Sequence 57297, A

## ALIGNMENTS

RESULT 1					
US-11-117-187-203					
; Sequence 203, Application US/11117187					
; Publication No. US2005026560A1					
; GENERAL INFORMATION:					
; APPLICANT: PREUS, DAPHNE					
; APPLICANT: COPEHAEVER, GREGORY					
; TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS					
; FILE REFERENCE: ARCD:309US					
; CURRENT APPLICATION NUMBER: US/11/117,187					
; CURRENT FILING DATE: 2005-04-28					
; PRIOR APPLICATION NUMBER: US/09/531,120					
; PRIOR FILING DATE: 2000-03-17					
; PRIOR APPLICATION NUMBER: 60/125,219					
; PRIOR FILING DATE: 1998-03-18					
; NUMBER OF SEQ ID NOS: 212					
; SOFTWARE: PatentIn Ver. 2.1					
; SEQ ID NO 203					
; LENGTH: 96583					
; TYPE: DNA					
; ORGANISM: Arabidopsis thaliana					
US-11-117-187-203					
Query Match					
Best Local Similarity 46.8%; Pred. No. 0.17;					
Matches 140; Conservative 0; Mismatches 159; Indels 0; Gaps 0;					
QY	263	GTTCCTCTCGACATCTTCACTCAATATCTATGTCTTAAACGACATTAATCAG	322		
DB	59989	GTGTGATTTCATGTTTGAGCACTATTTGTGATCTTATGATCATATATATAT	60048		
QY	322	CCAAACCACTTAAGTTATTTATAGTATTTTGTAAAGCACAGACATGTAGTG	382		
DB	60049	CAAAATTAATCTTAAGCACTCTTCAATTAAGAGTCAATTCAGCAATTAAGGATG	60108		
QY	383	CTGAGAAACGATGTGTGTAATCTTATTAATATCAAACTGGGTTAAATATAAAA	442		
DB	60109	AATTCACAAAGTTCGATTCACGCAATGATTAATGCAATTCAGTAAGCTAAGTAA	60168		
QY	443	AAATGTGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	502		
DB	60169	AATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	60228		
QY	503	ACGACATTTACCAAGGCTGATCTTGAAGCTTATTAATTAATTAATTAATTAATTA	561		
DB	60229	CAAGATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA	60287		

```

RESULT 2
US-10-750-185-25798/c
; Sequence 25798, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 25798
; LENGTH: 1711
; TYPE: DNA
; ORGANISM: Bovine 19866881051295
US-10-750-185-25798

Query Match
Best Local Similarity 46.3%; Score 42.4; DB 6; Length 1711;
Matches 139; Conservative 0; Mismatches 161; Indels 0; Gaps 0;

QY 260 TTGTTTCTCTGACATCTTCACCTCAATATCTTTGTTACTTAACAGCAATTAATC 319
DB 1063 TTCTCTAGCCTCAATATACAGATTGAGGAGGAGATATCAATTAATTTTAAAA 1004
QY 320 AGGCCAACCACTTAAAGTTTATTTGTATGATTTTGTGTAAGGACAGACATGTA 379
DB 1003 ATATATTTCCCTTTTCAATTAATTAATCTGCAATATATTAGCATGTCTCAATTTGA 944
QY 380 GTGCTGAGAAACTGATGTTGTAATCTTAATTAATATCAAACTGGTTAAATAA 439
DB 943 CTCAGGTGGAAGAACTGTGAATGATATCTCCAGAAAAATCATGATATTTCAGAGTAA 884
QY 440 AAAAAATGCACTAACTTAAAAAACAATACCAACAGACTTACTCCCTTGG 499
DB 883 AATAAATTCATGATGAGGATATGATTACAGAGAGAGGAAATATGCAATTTTGG 824
QY 500 AAAAGCACTTTACAAGGCTGATCTTAAAGCTTTATTTTACATTAAGAAATTAACA 559
DB 823 AAGAAAGTCTGTAGAAAAAGAGATATTTTAAAAACATTAACATGAGAAAAAAGCA 764

RESULT 3
US-10-750-185-55754/c
; Sequence 55754, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1

```

```

; SEQ ID NO 55754
; LENGTH: 1625
; TYPE: DNA
; ORGANISM: Bovine 19866881253827
US-10-750-185-55754

Query Match
Best Local Similarity 46.1%; Score 41.6; DB 6; Length 1625;
Matches 140; Conservative 0; Mismatches 164; Indels 0; Gaps 0;

QY 182 GGGACTAATTATATCAGGAACATGCCACAGATGAGCCCGAGAAACCTGTATACG 241
DB 452 GAGAAATTAATTAAGCTGTGATTTCTAATAGGACAGAAAGGTGTCTCCAGAAAGAAAG 393
QY 242 TGAGGAAAGGTGATGTTGTTTCTCTGACATCTTCACTCAATATCTATTGTTA 301
DB 392 GAAATTAAGATTAATCTGATTTCTGATGATCCAAACATTTATATGTTTATTTGTTA 333
QY 302 CTTAACAGCAATTAATCAGGCCAACCACTTAAGTTTATTTGTATATGTTTGTG 361
DB 332 GAGAAATGCAATTAATGAGCAAAAGTAAATTAATTTCAAACTCAGATATTTTGA 273
QY 362 TAAAGCAGACATGTGAGTGTGAGAAACTGATGTTGTTACTTGAATTAATATC 421
DB 272 GAAACATTAAGTTGATTAAGAAAGAAATGATCATATGTAATCTGTGACACAGCTG 213
QY 422 AAACGTGGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 481
DB 212 CAAAAAATTTGCAATGATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 153
QY 482 GACT 485
DB 152 AACT 149

RESULT 4
US-10-750-185-54114/c
; Sequence 54114, Application US/10750185
; Publication No. US20050260603A1
; GENERAL INFORMATION:
; APPLICANT: MMI GENOMICS, INC.
; APPLICANT: DENISE, Sue K.
; APPLICANT: KERR, Richard
; APPLICANT: ROSENFELD, David
; APPLICANT: HOLM, Tom
; APPLICANT: BATES, Stephen
; APPLICANT: FANTIN, Dennis
; TITLE OF INVENTION: COMPOSITIONS FOR INFERRING BOVINE TRAITS
; FILE REFERENCE: MM1100-2
; CURRENT APPLICATION NUMBER: US/10/750,185
; PRIOR FILING DATE: 2003-12-31
; PRIOR FILING DATE: 2002-12-31
; NUMBER OF SEQ ID NOS: 64922
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 54114
; LENGTH: 1593
; TYPE: DNA
; ORGANISM: Bovine 19866881235836
US-10-750-185-54114

Query Match
Best Local Similarity 58.3%; Score 40; DB 6; Length 1593;
Matches 70; Conservative 0; Mismatches 50; Indels 0; Gaps 0;

QY 372 ACATGTAGTGTGAGAAACATGATGTTGTAATCTGATTAATATATCAACAGCGTT 431
DB 1574 AAAATGATTTGATTAATTAATGATGATCTTTAACTTTTTCATGCTTTAAGAAATTTGAT 1515
QY 432 AAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 491
DB 1514 ATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTT 1455

```

[illegible]



Query Match	3.1%	Score 37.8	DB 6	length 3631
Best Local Similarity	48.4%	Pred. No. 2		
Matches 105	Conservative 0	Mismatches 112	Indels 0	Gaps 0
QY	407	TCATTATATATATCAAACTGGGTTAAAAAATGTCATTACTTAAAAAAA	466	
Db	2102	TGGTGATATCATTTTGATATTTGGCAAACTAATATCAATCATGTAAAGTTAAAAATATA	2167	

RESULT 14  
US-11-117-187-211/c  
: Sequence 211, Application US/11117187  
: Publication No. US2005026560A1  
: GENERAL INFORMATION:  
: APPLICANT: PREUS, DAPHNE  
: APPLICANT: COPEHAEVER, GREGORY  
: TITLE OF INVENTION: PLANT ARTIFICIAL CHROMOSOME COMPOSITIONS AND METHODS  
: FILE REFERENCE: ARCO:309US  
: CURRENT APPLICATION NUMBER: US/11/117,187  
: CURRENT FILING DATE: 2005-04-28  
: PRIOR APPLICATION NUMBER: US/09/531,120



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Comugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 12, 2005, 11:41:17 ; Search time 6216 Seconds

(without alignments)  
11156.533 Million cell updates/sec

Title: US-10-627-075-1

Perfect score: 1220  
Sequence: 1 gggcttccccagaccgacac.....ccaaaagagagagagatgc 1220

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 5083141 seqs, 28421725653 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenBank: \*  
1: gb\_ba: \*  
2: gb\_in: \*  
3: gb\_env: \*  
4: gb\_cm: \*  
5: gb\_ov: \*  
6: gb\_pat: \*  
7: gb\_ph: \*  
8: gb\_pr: \*  
9: gb\_ro: \*  
10: gb\_sbs: \*  
11: gb\_sy: \*  
12: gb\_un: \*  
13: gb\_vl: \*  
14: gb\_hg: \*  
15: gb\_pl: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1218.4	99.9	228046	9 AC129567	AC129567 Mus muscu
2	793.4	65.0	228774	14 AC114046	AC114046 Rattus no
3	793.4	65.0	326993	14 AC132960	AC132960 Rattus no
4	243.4	20.0	191377	8 AC018398	AC018398 Homo sapi
5	239	19.6	183215	14 AC091095	AC091095 Homo sapi
6	238	19.5	240825	6 AX087869	AX087869 Sequence
7	238	19.5	240825	6 AX523960	AX523960 Sequence
8	237.4	19.5	6714	8 AY563557	AY563557 Homo sapi
9	236.8	19.4	2471	9 BC027216	BC027216 Mus muscu
10	213	17.5	2424	9 AF004326	AF004326 Mus muscu
11	190.6	15.6	221218	14 AC152689	AC152689 Bos tauru
12	176.2	14.4	290	8 HS16468F	HS16468F Bos tauru
13	146.92	12.0	132392	14 AC141003	AC141003 Rattus no
14	143.4	11.8	763	8 HSA289781	HSA289781 Homo sapi
15	140.6	11.5	2846	6 AX458641	AX458641 Sequence
16	139.4	11.4	2846	6 AX458642	AX458642 Sequence
17	58	4.8	2000	6 AX655393	AX655393 Sequence
18	57.6	4.7	13925	2 AY591323	AY591323 Steinern

19	55.2	4.5	205066	14 AC162359	AC162359 Bos tauru
20	35	4.5	10710	6 AX345795	AX345795 Sequence
21	54.6	4.5	163149	14 CR855390	CR855390 Danio rer
22	54.4	4.5	15698	6 AX347043	AX347043 Sequence
23	54.4	4.5	110000	14 PFMAL13_25	PFMAL13_25 o
24	54	4.4	176966	5 CR394545	CR394545 Zebrafish
25	53.8	4.4	110000	2 AC116957_0	AC116957 Dictyoste
26	53.8	4.4	110000	2 AC116957_1	AC116957 Dictyoste
27	53.4	4.4	1084	2 AB101021	AB101021 Carabus m
28	53.4	4.4	1084	2 AB101022	AB101022 Carabus m
29	53.2	4.4	231489	14 AC156865	AC156865 Bos tauru
30	53	4.3	125958	2 AC115592	AC115592 Dictyoste
31	53	4.3	159184	14 BX36345	BX36345 Danio rer
32	53	4.3	237303	14 CR847960	CR847960 Danio rer
33	52.8	4.3	236582	14 CR388102	CR388102 Danio rer
34	52.6	4.3	1084	2 AB101024	AB101024 Carabus m
35	52.4	4.3	182871	2 AC117176	AC117176 Dictyoste
36	52.2	4.3	166384	5 BX294132	BX294132 Zebrafish
37	52.2	4.3	285836	14 AC128556	AC128556 Rattus no
38	52	4.3	226607	14 CR847902	CR847902 Danio rer
39	51.8	4.2	1522	2 AY701231	AY701231 Orconete
40	51.6	4.2	2606	15 YSCMTG07	YSCMTG07 SaccharomyC
41	51.6	4.2	6523	6 CO806912	CO806912 Sequence
42	51.6	4.2	237303	6 CO807186	CO807186 Sequence
43	51.6	4.2	6523	6 AX344713	AX344713 Sequence
44	51.6	4.2	6523	6 AX767488	AX767488 Sequence
45	51.6	4.2	6523	6 AX767564	AX767564 Sequence

## ALIGNMENTS

RESULT 1  
AC129567 228046 bp DNA linear ROD 06-APR-2004  
LOCUS AC129567  
DEFINITION Mus musculus chromosome 8, clone RP23-50L10, complete sequence.  
ACCESSION AC129567  
VERSION AC129567.8 GI:46240921  
KEYWORDS  
SOURCE  
MUS musculus (house mouse)  
ORGANISM

REFERENCE  
AUTHORS Birren, B., Nusbaum, C., Lander, E., Allen, N., Anderson, S., Barta, N., Bastien, V., Bloom, T., Boguslavsky, L., Boukhgalter, B., Camarata, J., Chang, J., Chazaro, B., Choquet, Y., Collamore, A., Cook, A., Cooke, P., DeArillano, K., Dewar, K., Diaz, J.S., Dodge, S., Fero, S., Ferreira, P., Fitzgerald, M., Gage, D., Galagan, J., Gardyna, S., Gord, S., Graham, L., Grand-Pierre, N., Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C., Kamat, A., Karalas, A., Kells, C., Landers, T., Levine, R., Lindblad-Toh, K., Liu, G., Maclean, C., MacDonald, P., Major, D., Matthews, C., McCarthy, M., Meldrum, J., Menuez, L., Mihova, T., Mlenga, V., Murphy, T., Navjot, J., Nguyen, C., Nicol, R., Norbu, C., Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J., Peterson, K., Phunhahng, P., Pierre, N., Raymond, C., Retta, R., Rise, C., Rogov, P., Roman, J., Roy, A., Schauer, S., Schupbach, R., Seaman, S., Severy, P., Smith, C., Spencer, B., Stange-Thomann, N., Stojanovic, N., Talamas, J., Tesfaye, S., Theodore, J., Topham, K., Travers, M., Vasilev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Young, G., Zainoun, J., Zembek, J., Zimmer, A. and Zody, W.  
DIRECT SUBMISSION  
Submitted (30-JUL-2002) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 228046)  
REFERENCE  
AUTHORS Birren, B., Nusbaum, C., Lander, E., Abouelleil, A., Allen, N., Anderson, M., Arachchi, H.M., Barta, N., Bastien, V., Bloom, T.,





repeat\_region /rpt\_family="TA)n"  
15535..15602  
/rpt\_family="(CATATA)n"  
15604..15705  
repeat\_region /rpt\_family="PB1D7"  
15706..15772  
/rpt\_family="ID3"  
repeat\_region complement(15813..16133)  
/rpt\_family="MTB"  
repeat\_region complement(16623..16776)  
/rpt\_family="Charlies"  
17613..17728  
repeat\_region /rpt\_family="PB1D10"  
18154..18320  
/rpt\_family="B3A"

Query Match 99.9%; Score 1218.4; DB 9; Length 228046;  
Best Local Similarity 99.9%; Pred. No. 1.8e-270;  
Matches 1219; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

1 GAGCTTCCAGTACCGATCTCTGACATTTAATTCTAGTCATGAAAGGGGTGACTC 60  
165469 GAGTCTTCCAGTACCGATCTCTGACATTTAATTCTAGTCATGAAAGGGGTGACTC 165528  
61 TGGACAGAGAGCCACAGAGCTGGAAGCTTTAAGTCAGTCAGCCCGACCTTTT 120  
165529 TGGACAGAGAGCCACAGAGCTGGAAGCTTTAAGTCAGTCAGCCCGACCTTTT 165588  
121 ATGGCAGAGGGCTTTTGAATTAATTAAGGGAAGTATTTGCTGAGCCCACTGAC 180  
165589 ATGGCAGAGGGCTTTTGAATTAATTAAGGGAAGTATTTGCTGAGCCCACTGAC 165648  
181 TGGGACTAATTTATCAGGAACATGCGACAGAGTATGAGCCCGAGAAAACCTGATACA 240  
165649 TGGGACTAATTTATCAGGAACATGCGACAGAGTATGAGCCCGAGAAAACCTGATACA 165708  
241 GGAAGGAAGAGGTATGTTTCTTCCCTGACATCTTCACTCAATATCTATTTT 300  
165709 GGAAGGAAGAGGTATGTTTCTTCCCTGACATCTTCACTCAATATCTATTTT 165768  
301 ACTTAACAGACATTAATATCAGGCCAACAACCTTAATTTTATATATTTT 360  
165769 ACTTAACAGACATTAATATCAGGCCAACAACCTTAATTTTATATATTTT 165828  
361 TTAAGGCACAGATGATGCTGAGAAAACGATGTTGTTAATTTAATTAAT 420  
165829 TTAAGGCACAGATGATGCTGAGAAAACGATGTTGTTAATTTAATTAAT 165888  
421 CAAACCTGGTTAAAAATGTCATTAATTAATTAATTAATTAATTAATTAAT 480  
165889 CAAACCTGGTTAAAAATGTCATTAATTAATTAATTAATTAATTAATTAAT 165948  
481 AGACTTTACTCCCTTGGAAAAAGACATTTTACAAAGGCTTATAGCTTTATTTA 540  
165949 AGACTTTACTCCCTTGGAAAAAGACATTTTACAAAGGCTTATAGCTTTATTTA 166008  
541 CAATAAAGAAATTAACCAAGGTCCGATATAGCTTAAATTTTCTTAAGAACAG 600  
166009 CAATAAAGAAATTAACCAAGGTCCGATATAGCTTAAATTTTCTTAAGAACAG 166068  
601 AACTTTACTATGCTTTAAATTAAGTATTAATCTCAATCTCTGCAAGCTTACCTA 660  
166069 AACTTTACTATGCTTTAAATTAAGTATTAATCTCAATCTCTGCAAGCTTACCTA 166128  
661 CAAACGAGCAGACAGACAGAGCCCGACCTACTCTAGGAAATTAATTAAGGCTGTC 720  
166129 CAAACGAGCAGACAGACAGAGCCCGACCTACTCTAGGAAATTAATTAAGGCTGTC 166188  
721 CTCTGACATGCCAGGGGCTTGTGCTGCTGTCTGTCTCCAGAAAGCTTTCTGACATACA 780  
166189 CTCTGACATGCCAGGGGCTTGTGCTGCTGTCTGTCTCCAGAAAGCTTTCTGACATACA 166248  
781 CAGTCTTTGGGGCAGTACGACTATGCTGATTTTCTGTGCTGCTGCTGACATGACCC 840

DB 166249 CAGTCTTTGGGGCAGTACGACTATGCTGCTGATTTTTTCTGTGCTGCTGACATGACCC 166308  
QY 841 CCTACAGAAATAGTGGGTGAGCCAGGGGGGAGAGCGGCTGCTGACATGCTGCTG 900  
DB 166309 CCTACAGAAATAGTGGGTGAGCCAGGGGGGAGAGCGGCTGCTGACATGCTGCTG 166368  
QY 901 CTCTTATCACTTATATTAAGGGAAGAAAGTATTAATTCGATCTGACACTGTAG 960  
DB 166369 CTCTTATCACTTATATTAAGGGAAGAAAGTATTAATTCGATCTGACACTGTAG 166428  
QY 961 ACTCAGGGGAGAAACAAAGATCCCGACAGACCTCTGAGTAGAGAGGGCTGCTTCC 1020  
DB 166429 ACTCAGGGGAGAAACAAAGATCCCGACAGACCTCTGAGTAGAGAGGGCTGCTTCC 166488  
QY 1021 TCTCAGAGAGCTCCGAGTGTCCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080  
DB 166489 TCTCAGAGAGAGCTCCGAGTGTCCGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 166548  
QY 1081 GAGCTGTCTGGGAGCGAGAAAGCTCTCACTGATGACTTATTCACAGGACAGCCCT 1140  
DB 166549 GAGCTGTCTGGGAGCGAGAAAGCTCTCACTGATGACTTATTCACAGGACAGCCCT 166608  
QY 1141 GTGCTTACAGACAGCTGAGAGCTCAGAGCCCAAGTTTGTAATCACTGATTTAGAC 1200  
DB 166609 GTGCTTACAGACAGCTGAGAGCTCAGAGCCCAAGTTTGTAATCACTGATTTAGAC 166668  
QY 1201 CCAAAAAGAGAGAGAGATG 1220  
DB 166669 CCAAAAAGAGAGAGAGATG 166688

RESULT 2  
AC114046 228774 bp DNA linear HTG 13-MAY-2003  
LOCUS Rattus norvegicus clone CH230-18708, WORKING DRAFT SEQUENCE.  
DEFINITION AC114046  
AC114046 GI:30581607  
VERSION HTG; HTGS\_PHASE2; HTGS\_DRAFT; HTGS\_FULLTOP.  
KEYWORDS Rattus norvegicus (Norway rat)  
SOURCE Rattus norvegicus  
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
1 (bases 1 to 228774)  
Muzny,D,Marie, Metzker,M,Lee, Adamson,S, Adams,C, Alder,J, Allen,C, Allen,H, Alsbrooks,S, Amin,A, Anguiano,D, Anyalebechi,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H, Baldwin,D, Bandaranaike,D, Barber,M, Barnstead,M, Benhmed,F, Biswal,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M, Bryant,N, Buhay,C, Burch,P, Butrell,K, Calderon,E, Cardenas,V, Carter,K, Cavazos,I, Caesar,H, Center,A, Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J, Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, P'Souza,L, Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D, Delgado,O, Denison,S, Deramo,C, Ding,Y, Dimh,H, Divya,K, Draper,H, Dugan-Rocha,S, Dunn,A, Durbin,K, Duval,B, Eaves,K, Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G, Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P, Frazer,C,M, Gabisi,A, Ganta,R, Garcia,A, Garner,T, Garza,M, Gebregeorgis,B, Geet,K, Gill,R, Grady,M, Guerra,M, Guevara,M, Gunaratne,P, Haaland,W, Hamill,C, Hamilton,C, Hamilton,K, Harvey,Y, Havlik,P, Hawes,A, Henderson,N, Hernandez,J, Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hogues,M, Hollins,B, Howell,S, Huijck,S, Hume,J, Idlerbird,D, Jackson,A, Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jolyet,A, Karpathy,S, Kelly,S, Kelly,S, Khan,Z, King,L, Kovar,C, Kowis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J, Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J, Lorenshuew,L, Louisedge,H, Lozada,R,J, Lu,X, Ma,J, Maheshwari,M, Mahindaratne,M, Mahmoud,M, Malloy,K, Mangum,B, Mapua,P, Martin,K, Martin,R, Martinez,E, Mawhinney,S, McLeod,M,P, McNeill,T,Z, Meenen,B,

Milosevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Murtida, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwankwelu, O., Okunolu, G., Olarnpusagoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfannkuch, C., Plopper, F., Polidexter, A., Popovic, D., Primus, E., Pu, L., Puzos, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M., Reigh, R., Riley, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Sanders, M., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, U., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smajs, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soosa, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Usmani, K., Valas, R., Vera, V., Villanar, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczky, R., Wood, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, L., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhou, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G. and Gibbs, R.A.

Direct Submission  
Unpublished  
2 (bases 1 to 228774)  
Worley, K.C.

Direct Submission  
Submitted (07-MAR-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
3 (bases 1 to 228774)  
Rat Genome Sequencing Consortium.

Direct Submission  
Submitted (13-MAY-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
On May 13, 2003 this sequence version replaced gi:22857247. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rat/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: CH230-187J8  
----- Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 215732 bases at least Q40  
Consensus quality: 217275 bases at least Q30  
Consensus quality: 218412 bases at least Q20  
Estimated insert size: 225110; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

\* NOTE: Estimated insert size may differ from sequence length  
(see [http://www.hgsc.bcm.tmc.edu/docs/Gendbmk\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/Gendbmk_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 1 contigs. Gaps between the contigs  
\* are represented as runs of N. The order of the pieces  
\* is believed to be correct as given, however the sizes  
\* of the gaps between them are based on estimates that have  
\* provided by the submitter.  
\* This sequence will be replaced

\* by the finished sequence as soon as it is available and  
\* the accession number will be preserved.  
1 228774: contig of 228774 bp in length.  
Location/Qualifiers  
source  
1. 228774  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-187J8"  
1. 1046  
misc\_feature  
/note="wgs contig"  
227275..228774  
misc\_feature  
/note="wgs\_contig"

ORIGIN  
Query Match 65.0%; Score 793.4; DB 14; Length 228774;  
Best Local Similarity 85.0%; Pred. No. 2,1e-172;  
Matches 1079; Conservative 0; Mismatches 111; Indels 79; Gaps 14;

2 AGCTTCCAGTACGATCTCTGAGCATTTAATTCTAGTCATGAAAGGGTGTGACTCT 61  
|||||  
82805 AGCTGCAACAGTACCAATCTCTGAGCATTTAATTCTAGTCATGAAAGGGTGTGACT-- 82862  
|||||  
62 GGACGACGAGGCCACAGCTGGAAGTCT--TTTGAAGTCAGTCCAGCCAGCTTTT 120  
|||||  
82863 ---GGACGACGAGGCCACAGCTGGAAGTCTTTTGAAGTCAGTACCCCGACCTTTT 82919  
|||||  
121 ATGGCAGGGGCTTTTGAATTTAATAAGGGGAATGATTTGCTGAGCCGCTGAC 180  
|||||  
82920 ATGGCCAAAGGGTTTGAATTTAATAAGGGGAATGATTTGCTGAGCTGACTTATAC 82979  
|||||  
181 TGGGCTAATTTATACGAAATGCGACAGATGATGAG--CCGAGGAAACCTGATAC 239  
|||||  
82980 TGAGCTAATTTATACGAAATGCGACAGATGATGAG--CCGAGGAAACCTGATAC 83039  
|||||  
240 AGTGAAGGAAAGGATGTTGTTGTTTCTCTGACATGATCTCACTCAAAATATCATTTGT 299  
|||||  
83040 AGCGAAGGAAAGGATGTTGTTGTTTCTCTGACATGATCTCACTCAAAATATCATTTGT 83095  
|||||  
300 TACTTACAGACAAATTAATGAGCCAAACCACTTAATTTATTTATTTATTTTGT 359  
|||||  
83096 TACTTACAGACAAATTAATGAGCCAAACCACTTAATTTATTTATTTATTTTGT 83155  
|||||  
360 G-----TTAAGGACAGACATGATGCTGAGAAACATGATGTTGTTATTTGA 409  
|||||  
83156 GTATTCTGTGTTAAAGGTTAAACATGATGCTGAG--AAGCTGACCTGATTAATTTGA 83213  
|||||  
410 TTTAATATATCAACTGAGGTTTAAAT---AAAAAAATGTGATTAATTAATAAAAA 466  
|||||  
83214 TTTAATATATCAACTGAGGTTTAAATTTAATAATAATAATAATAATAATAATAATA 83272  
|||||  
467 ACCAAATTCACAAAGACTTACTTCCCTTGAAACCACTTTACAAAGGCTGATCTT 526  
|||||  
83273 ---AAAAACAGCAACTTACTTCCCTTGAAACCACTTTAAGGCGGATCTT 83329  
|||||  
527 AGCCTTATATTTCAATTAAGAAATTAACCAAGTCCGATATAGTGTATTTATTT 586  
|||||  
83330 ACTCTTATATTTCAATTAAGAAATTAACCA--GTCCTAATTTACTGATGATTTATTT 83387  
|||||  
587 CCT-----AAAAAGACAGAACTTTCACTATGTTTAAATTT 623  
|||||  
83388 CTTAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAAAC 83447  
|||||  
624 AAAGTATTTACTCTGAGTACTCTGCAAGCTTAAAGCTTAAACGACGACGACGACG 683  
|||||  
83448 ACAAGTATTTAAACGAGTACTCTGCAAGCTTAAAGCTTAAACGACGACGACGACG 83507  
|||||  
684 GCCCAGCTACTCTCTGAGAAATTAATTAAGGTTGCTGCTGACATGCGCAGAGGCTTTG 743  
|||||  
83508 GCCCAGCTACTCTCTGAGAAATTAATTAAGGTTGCTGCTGACATGCGCAGAGGCTTTG 83567  
|||||  
744 TGAGTGTGTGTGTTCCAGAAAGCTTCTGACATACAGTCTCTTGGGAGTAAAGCAC 803  
|||||

Db	83568	TGGCTGCGTGTGTATCCAGAAAGGTTCTGCAGTGCAACAGTCCTTTGGGGCAGTAAGCAC	83628
Qy	804	TATGCTCATGATTTTTTCCCTGTTGCTGCTGCTAGTGACACCCCTTACAGAGAATATGCGGTGAG	863
Db	83628	TATGCTCATGATTTTTTCCCTGTTGCTGCTGCTAGTGACACCCCTTACAGAGAACAATGCGGTGAG	83687
Qy	864	CCAGGGGGCCGAGCGGCTGGGCTGCACATGTCGTGGCGCTCTTATGACAATTATCATATTAG	923
Db	83688	CCAGGGGGCCGAGCGGCTGGGCTGCACATGTCGTGGCGCTCTTATGACAATTATCATATTAG	83747
Qy	924	GGAAGAAAAGTGAATTGATTTCCGATACTGACACACTGATGACTCAGGGGAGAACAAAGAGTC	983
Db	83748	GGAAGAAAAGTGAATTGATTTCCGATACTGACACACTGATGACTCAGGGGAGAACAAAGAGTC	83807
Db	984	CGTGCAAGACCTCTGGAGTGAGCAGCGGGCTGCTCTTCTCTTCAGACAGCTCCGAGTGTGC	1043
Db	83808	CGTGCAAGACCTCTGGAGTGAGCAGCGGGCTGCTCTTCTCTTCAGACAGCTCCCTGTGTGC	83867
Qy	1044	CGGGGAGAAAGAGAGAAAGACACAGGCACTGGGAAAGACCTGCTCGGGGACGGAGAG	1103
Db	83868	CGG-----GAGAAGAGAAAGACCTGGGAAATAGCCTGCTTAG----GGACGGGAGAG	83917
Qy	1104	GCTCTCACTGATGAGACTTATTCACACGCGACACCCTGTGCTTGTAGACAGCACTGAGAG	1163
Db	83918	GCTCTCACTGATGAGACTTGTGCCA--GCACAGCCCTGTGCTTGTAGACAGCACTGAGAG	83975
Qy	1164	CTCAGGAC-----GCAGTTTGTCTGAACCTACACAGTTTGAACCCAAAAAGAGA	1211
Db	83976	CTCAGGACCTCTATGTGCTCTGCAAGTTTGCTGMACTCACAGTTTGAACCCAAAGAGAGA	84035
Qy	1212	GAGAGAAATG 1220	
Db	84036	AAGAATGTG 84044	
RESULT 3			
AC132960			
LOCUS			
DEFINITION	AC132960	326993 bp	DNA linear HTG 19-NOV-2002
VERSION	Rattus norvegicus clone CH230-162u3,	*** SEQUENCING IN PROGRESS	
KEYWORDS	***, 16 unordered pieces.		
SOURCE	AC132960.3 GI:25073326		
ORGANISM	HMG; HTGS_PHASE1; HTGS_DRAT; HTGS_ENRICHED. Rattus norvegicus (Norway rat)		
REFERENCE			
AUTHORS	Muzny,D.,Marle,M.,Metzger,M.,Lee,A.,Abramzon,S.,Adams,C.,Alder,J.,Allen,C.,Allen,H.,Albrooks,S.,Amn,A.,Angilano,D.,Angelabechi,V.,Aoyagi,A.,Ayodeji,M.,Baca,E.,Baden,H.,Balwin,D.,Bandaraike,D.,Barber,M.,Barnstead,M.,Benamed,F.,Blawie,K.,Blair,J.,Blankenburg,K.,Blyth,P.,Brown,M.,Byran,N.,Buhaý,C.,Burck,P.,Butrell,K.,Calderon,E.,Cardenas,V.,Cartier,K.,Cavazos,I.,Ceasar,H.,Center,A.,Chacko,J.,Chavez,D.,Chen,G.,Chen,R.,Chen,Y.,Chen,Z.,Chu,J.,Cleveland,C.,Cockrell,R.,Cox,C.,Coyle,M.,Crease,A.,D'Souza,L.,Davila,M.L.,Davis,C.,Davy-Carroll,L.,De Anda,C.,Dederich,D.,Delgado,O.,Denson,S.,Deramo,C.,Ding,Y.,Dinh,H.,Divya,K.,Diaper,H.,Dugan-Rocha,S.,Dunn,A.,Durbin,K.,Duval,B.,Eaves,K.,Egan,A.,Escotto,M.,Eugene,C.,Evans,C.A.,Falls,T.,Fan,G.,Fernandez,S.,Finley,M.,Flanagan,N.,Forbes,L.,Foster,M.,Foster,P.,Fraser,C.M.,Gabisi,A.,Ganta,R.,Garcia,A.,Garner,T.,Garrza,W.,Georgescu,E.,Geer,K.,Gill,R.,Grady,M.,Guerra,W.,Gunaratne,P.,Haaland,W.,Hamil,C.,Hamilton,C.,Hamilton,K.,Harvey,Y.,Hevlik,P.,Hawes,A.,Henderson,N.,Hernandez,U.,Hernandez,R.,Hines,S.,Hladun,S.L.,Hodgson,A.,Hogues,M.,Hollins,B.,Howell,S.,Huylk,S.,Hume,J.,Idlbird,D.,Jackson,A.,Jacksen,L.,Jacob,L.,Jiang,H.,Johnson,B.,Johnson,R.,Jolivet,A.,Karpachy,S.,Kelly,S.,Kelly,S.,Khan,Z.,King,L.,Kovacs,C.,Kowis,C.,Kraft,C.L.,Lebow,H.,Levan,J.,Lewis,L.,Li,Z.,Liu,J.,Liu,W.,Liu,M.,Liu,Y.,London,P.,Longacre,S.B.,Lopez,J.J.,		

[illegible]

\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```
1 22617: contig of 22617 bp in length
* 22618 22717: gap of unknown length
* 22718 36669: contig of 13952 bp in length
* 36670 36769: gap of unknown length
* 36770 41082: contig of 4313 bp in length
* 41083 41182: gap of unknown length
* 41183 47337: contig of 6155 bp in length
* 47338 47437: gap of unknown length
* 47438 51040: contig of 3603 bp in length
* 51041 51140: gap of unknown length
* 51141 278324: contig of 227184 bp in length
* 278325 312760: contig of 34336 bp in length
* 312761 312860: gap of unknown length
* 312861 313933: contig of 1073 bp in length
* 313934 314033: gap of unknown length
* 314034 315121: contig of 1088 bp in length
* 315122 315221: gap of unknown length
* 315222 316822: contig of 1401 bp in length
* 316823 317808: gap of unknown length
* 317809 317908: contig of 1086 bp in length
* 317909 319328: gap of unknown length
* 319329 319428: gap of 1420 bp in length
* 319429 320859: contig of 1431 bp in length
* 320860 320959: gap of unknown length
* 320960 322300: contig of 1341 bp in length
* 322301 322400: gap of unknown length
* 322401 323510: contig of 1110 bp in length
* 323511 326993: contig of 3383 bp in length.
* 323611 Location/Qualifiers
```

FEATURES  
source  
1. 326993  
/organism="Rattus norvegicus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:10116"  
/clone="CH230-162J3"

```
misc_feature  
1. 1068  
/note="wgs_end_extension  
clone_end:5p6"  
5479..5887  
/note="clone_boundary  
clone_end:5p6  
site:  
end_sequence:BH270833"  
16123..17806  
/note="wgs_contig"  
22618..22717  
/estimated_length=unknown  
misc_feature  
22718..23793  
/note="wgs_contig"  
35557..36669  
/note="wgs_contig"  
36670..36769  
/estimated_length=unknown  
misc_feature  
36770..37959  
/note="wgs_contig"  
41083..41182  
/estimated_length=unknown  
misc_feature  
41183..42476  
/note="wgs_contig"  
47338..47437  
/estimated_length=unknown  
misc_feature  
47438..49026  
/note="wgs_contig"  
51041..51140  
/estimated_length=unknown  
gap
```

```
misc_feature 51141..53764  
/note="wgs_contig"  
misc_feature 53815..54829  
/note="wgs_contig"  
misc_feature 54880..55894  
/note="wgs_contig"  
misc_feature complement(223969..224620)  
/note="clone_boundary  
clone_end:T7  
site:  
end_sequence:BH270832"  
misc_feature 277235..278324  
/note="wgs_end_extension  
clone_end:T7"  
gap 278325..278424  
/estimated_length=unknown  
misc_feature 278425..279604  
/note="wgs_end_extension
```

Query Match 65.0%; Score 793.4; DB 14; Length 326993;  
Best Local Similarity 85.0%; Pred. No. 2e-172;  
Matches 1079; Conservative 0; Mismatches 111; Indels 79; Gaps 14;

```
2 AGCTTCCAGTACGATCTCTGACGACATTAATCTTGTGATGATGAGGGGTGACCT 61  
Db 227256 AGCTTGCACAGTACCAATCTCTGAGCATTTAATCTTGTGATGAGGGGTGACT-- 227313  
Qy 62 GGACGACGAGCAGCAGAGCTGGAAGTGT--TTGAAATCAGTCCAGCCCCAGCTTTT 120  
Db 227314 ---GGACGACGAGCAGAGGCTGGAAGTGTCTTTAGAAAGTGTACGCCACCTTTT 227370  
Qy 121 ATGGCAGAGGGCTTTGAACTTAATTAAGGGGAAAGTATTTGCTGAGCCCATGAC 180  
Db 227371 ATGGCAGAGGGCTTTGAACTTAATTAAGGGGAAAGTATTTGCTGAGCTCATTTGAC 227430  
Qy 181 TGGGACTAATTTAATCAGAAACATGCAAGATGATGAG--CCGAGGAAACCCGATATAC 239  
Db 227431 TGAGCTAATTTAATCAGTAACATGCAAGATGATGAGCCCGAGGAAACCCGATATAC 227490  
Qy 240 AGTGAAGGAAAGGTGTATGTTGTTGTTCTCTGACATATCTTCAATATATCTTATGT 299  
Db 227491 AGCGAAGGAAAGGATATA---TGTTTCTCTGTGATATACCTCAATATATCTTATGT 227546  
Qy 300 TACTTAACGACATTAATACGCGCAACCACTTAATGTTTATTTGATATGATTTTGT 359  
Db 227547 TACTTAACGACATTAATACGCGCAACCACTTAATGTTTATTTGATATTTAAATGG 227606  
Qy 360 G-----TTAAGCAGACAGATGAGTGTGAGAAACATGATGTGTGATATCTTGA 409  
Db 227607 GTATTCGTGTGTTAAAGTAAACATGTAAGTGTGAG--AAGCTGACTGATTAATTGA 227664  
Qy 410 TTTATTAATATCAATCTGGTTAAAT--AAAAAAATGTGATTAATTAATAAAAAA 466  
Db 227665 TTTATTAATATCAATCTGGTTAAATTTAAATAAAAAATGATTAATCTAGTAGTA- 227723  
Qy 467 ACCAAATACCAACAAGATTAATCTCCCTTGGAAGCAATTTAACAAGGCTGATCTT 526  
Db 227724 ---AAAAACAGCAGAACTTAATCTCTTGGAAGCAACAATTTTAAGGCGATCTT 227780  
Qy 527 AGCCTTAATTTAATCAATTAAGAAATTAACAAGATGCTCCGATATGCTGATTTTATTT 586  
Db 227781 ACTCTTAATTTAATCAATTAAGAAATTAACA--GTCTTAATATAGCTGATGATTTATTT 227838  
Qy 587 CCT-----AAAAACAAGAACTTTTCACTATGCTTTTAAATTT 623  
Db 227839 CCTAAAAAACAACAACAACAACAACAACAACAACAACAACAACAACAACAACAATTT 227898  
Qy 624 AAAGTATTACCTCAGATCTCTGCAAGCTTAAGCTCAACGAGCAGACGACAAACGA 683  
Db 227899 AAGTATTACCTCAGATCTCTGCAAGCTTAAGCTCAACGAGCAGACGACAAACGA 227958  
Qy 684 GCCCAGTACTCTCTAGGAATTAATTAAGGTGTGCTCTGACATGCCAGGAGGCTTTG 743
```

```

Db      227959  GCCCAGACTTCTCTAGGAAATATTAAGGGTGTGCTTATGACATGCCAGGGGCTTGG 228018
QY      744      TGGCTGCTGTGTGTCTCCGAGAGGCTTTGCGAGTACAGACTCTTTGGGGCAGTAAAGC 803
Db      228019  TGGCTGTGTGTGTCTCCGAGAGGCTTTGCGAGTACAGACTCTTTGGGGCAGTAAAGC 228078
QY      804      TATGCTGTATTTTCTGTTGCTGTGCTAGTACACCCCTACAGAGAAATAGTGGGTGAG 863
Db      228079  TATGCTGTATTTTCTGTTGCTGTGCTAGTACACCCCTACAGAGAAATAGTGGGTGAG 228138
QY      864      CCAGGGGGGCGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 923
Db      228139  CCAGGGGGGCGAGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 228198
QY      924      GGAAGAAAGTGTATTTATTTCTGTTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 983
Db      228199  GGAAGAAAGTGTATTTATTTCTGTTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 228258
QY      984      CGTGACAGCTCTGCGAGTACAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1043
Db      228259  CGTGACAGCTCTGCGAGTACAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 228318
QY      1044      CGGAGGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1103
Db      228319  CGGAGGAGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 228368
QY      1104      GCTCTCACTGATGACTTATTTACAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1163
Db      228369  GCTCTCACTGATGACTTATTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 228426
QY      1164      CTCAGAGAC-----GCAGCTTCTGAACTCAGAGTTTAAACCCCAAAAGAGA 1211
Db      228427  CTCAGAGACCTCTATGTGCTCTGCAAGTTTCTGAACTCAGAGTTTAAACCCCAAAAGAGA 228486
QY      1212      GAGAGATG 1220
Db      228487  AAGAAATGTG 228495

RESULT 4
LOCUS      AC018398                      191377 bp    DNA    linear    PRI 30-APR-2002
DEFINITION Homo sapiens chromosome 8, clone RP11-16G12, complete sequence.
ACCESSION  AC018398
VERSION     AC018398.10  GI:20198704
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Homnidae; Homo.
REFERENCE   1 (bases 1 to 191377)
            Birren, B., Linton, L., Nusbaum, C., Lander, E., Abram, H., Allen, N.,
            Anderson, S., Baldwin, J., Barna, N., Beckerly, R., Beda, F.,
            Boguslavskiy, L., Boukhalter, B., Brown, A., Castle, A., Colangelo, M.,
            Collins, S., Collymore, A., Cooke, P., Dearellano, K., Dewar, K.,
            Domingo, M., Doyle, M., Feneester, J., Ferreira, P., Fitzhugh, W.,
            Forrest, C., Gage, D., Galagan, J., Gardyna, S., Grant, G., Hago, B.,
            Heath, A., Horton, L., Howland, J., Johnson, R., Jones, C., Kann, L.,
            Kartas, A., Klein, J., Landers, T., Lehoczy, J., Lieu, C., Locke, K.,
            Macdonald, P., Marquis, N., McEwan, P., McGurk, A., McKernan, K.,
            Meldrim, J., Morrow, J., Naylor, J., Norman, C. H., O'Connor, T.,
            O'Donnell, P., Peterson, K., Pierre, N., Pollara, V., Riley, R.,
            Rothman, D., Roy, A., Santos, R., Severy, P., Stange-Thomann, N.,
            Stojanovic, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
            Tittel, A., Vassiliev, H., Viel, R., Vo, A., Wu, X., Wyman, D., Ye, W., J.,
            Zimmer, A. and Zody, M.
            Direct Submission
            Submitted (10-DEC-1999) Whitehead Institute/MIT Center for Genome

```

```

REFERENCE   3 (bases 1 to 191377)
            Birren, B., Linton, L., Nusbaum, C., Lander, E., Ali, A., Allen, N.,
            Anderson, S., Barna, N., Bastien, V., Bloom, T., Boguslavskiy, L.,
            Boukhalter, B., Brown, A., Camarata, J., Campoliano, A., Chang, J.,
            Chazaro, B., Choepel, Y., Colangelo, M., Collins, S., Collymore, A.,
            Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S.,
            Fero, S., Ferreira, P., Fitzhugh, W., Gage, D., Galagan, J., Gardyna, S.,
            Garte, S., Gird, S., Goyette, M., Graham, L., Grand-Pierre, N.,
            Hago, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,
            Kamat, A., Kartas, A., Kells, C., LaRocque, K., Lamazares, R.,
            Landers, T., Lehoczy, J., Levine, R., Lindblad-Toh, K., Liu, G.,
            Maclean, C., Macdonald, P., Major, J., Marquis, N., Matthews, C.,
            McCarthy, M., McEwan, P., McKernan, K., Meldrim, J., Meneus, L.,
            Mithova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R.,
            Norbu, C., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neill, D.,
            Oliver, J., Peterson, K., Phunkhang, P., Pierre, N., Pollara, V.,
            Raymond, C., Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P.,
            Roman, J., Rosetti, M., Roy, A., Santos, R., Schauer, S., Schnupack, R.,
            Seaman, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
            Strauss, N., Subramanian, A., Talamas, J., Testaye, S., Theodore, J.,
            Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,
            Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W., J., Young, G.,
            Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.
            Direct Submission
            Submitted (19-APR-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
            4 (bases 1 to 191377)
TITLE
JOURNAL
REFERENCE
AUTHORS

```

```

TITLE
JOURNAL
COMMENT

```

## FEATURES

## source

```

-----
Location/Qualifiers
1..191377
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="8"

-----
Center: Whitehead Institute/ MIT Center for Genome Research
Center code: WIBR
Web site: http://www-seq.wi.mit.edu
Contact: sequence.submissions@genome.wi.mit.edu
-----
Project Information
Center project name: 13535
Center clone name: 16_G_12
-----

```

```
/clone="RP11-16G12"
/clone_lib="RP11-1 Human Male BAC"
repeat_region
/rpc_family="AluSg/x"
1..221
complement(1277..1428)
/rpc_family="L1ME1"
repeat_region
1732..1823
/rpc_family="(CA)n"
2184..2238
/note="PCR product sequence only"
complement(2649..2774)
/rpc_family="L2"
repeat_region
2963..3133
/rpc_family="(TG)n"
3167..3208
/rpc_family="(TG)n"
3581..3587
/note="<30 qual SNGL region"
complement(6732..6840)
/rpc_family="L2"
repeat_region
complement(8405..8540)
/rpc_family="MIR3"
8922..9027
/rpc_family="L2"
repeat_region
9061..9174
/rpc_family="L2"
9390..9595
/rpc_family="L2"
repeat_region
complement(9592..9699)
/rpc_family="MIR"
1109T..11333
/rpc_family="AluSx"
repeat_region
11359..11649
/rpc_family="AluSp"
11765..12100
/rpc_family="AluJb"
12325..12593
/rpc_family="AluJb"
13535..13813
/rpc_family="AluSp"
13817..13850
/rpc_family="(CAA)n"
13857..14036
/rpc_family="MTR"
17230..17345
/rpc_family="L1ME1"
17528..17551
/rpc_family="AT-rich"
1847T..18757
/rpc_family="AluJc"
21235..21353
/rpc_family="L2"
21420..21484
/rpc_family="L1PA3"
22294..22456
/rpc_family="MTR"
complement(22452..22513)
/rpc_family="MIR"
22668..23024
/rpc_family="MLT1A1"
23390..23520
/rpc_family="MER85"
23535..23739
/rpc_family="MER2"
23908..24013
/rpc_family="MIR"
24043..24073
/rpc_family="(TTTG)n"
complement(24079..24356)
/rpc_family="AluSg"
24877..24952
/rpc_family="LTR67"
27866..28203
repeat_region
```

```
repeat_region
/rpc_family="AluSx"
28205..28272
/rpc_family="GA-rich"
28663..28682
/rpc_family="(T)n"
complement(28688..29140)
/rpc_family="MLT1C"
repeat_region
29551..29760
/rpc_family="L2"
repeat_region
complement(29782..30036)
/rpc_family="L1ME4"
repeat_region
30200..30591
/rpc_family="MLT1B"
complement(30615..30910)
/rpc_family="L1ME4"
31381..31634
/rpc_family="L2"
31679..31699
/rpc_family="AT-rich"
31939..31990
/rpc_family="L2"
repeat_region
33132..33161

Query Match 20.0% Score 243.4; DB 8; Length 191377;
Best Local Similarity 64.2% Pred. No. 1.6e-45;
Matches 554; Conservative 0; Mismatches 271; Indels 38; Gaps 11;

QY 206 CCACAGAGTGAATGAGCCCGAGAAACCTGATACAGTGAAGAAAGGTGATGTTT 265
DB 40710 CGACAGAGCGAGACTCCGTCACAAAAAAGTTGATATTCCTCT 40769
QY 266 TCCTTCGACATCTTCACTCAATATCTATTGTTACTTAACAGCAATTAATCGGCCA 325
DB 40770 TCTACACAGGACCTTAATCTTAATTTCTTATGTAATGTAATGACTTAATCATGTC 40829
QY 326 AACCACTTAAGTTTATTTGATATGTAATTTGTTGAAGCAGACATGAGTGTG 385
DB 40830 GTGTATTTTGAAGGATATCCAGATGAGATTTTAAAGA-ATAAGGTCAAGCTC 40888
QY 386 AGAAAACTGATGTTGTAATCTGATTAAATA-----TATCAACTGGGTTAAA 435
DB 40889 AG-AAACTGATGTTCTTAACGAGCTTAATAACAGATCTTCTATCAACTGTTGGAAA 40947
QY 436 TAAAAA-----AAATGTGATATCTTAATAAAAAAACAATACCAACAGACTTTA 488
DB 40948 TGAATATCTTTGTAACTGTGCTAGTGTTCAGATCACTAAATCAATGAATCTAT 41007
QY 489 CTTCCTTGGAAAAAGCATTACAGAG--CTGATCTAGCCTTATATTTACATA 545
DB 41008 TTTCCCTCTGAAAAAGCAGACGTTTACAGAGCCAACTTCTCTCTTTATTTGTAATA 41067
QY 546 AAGAAATTAACCAAGGTCCGATATAGCTGAATTTATTCCTA-----AAGAAACA 598
DB 41068 ACAAATATTAACCGAAGTCTGACCTATTTGAATATTTATTCCTAAGAAAAACA 41127
QY 599 GAACTTTCACTATGCTTTAAATTAAGTGATTAACCTCAATACTCTGCAAGCTTACCC 658
DB 41128 GGAATCTTATGTAATCTTCAACATTAAGTAATTAACCTCAATATTTGCAAGCTTACCA 41187
QY 659 TACAACAGAGAGA--CAGACAAACAGAGCCAGCT-ACCTCTGAGAAATATTA----- 711
DB 41188 CGGCAAAAATCAGTTTCAAGACAAAGAGATCAACGCTCTCTCTGAGAAATATCTTAATTG 41247
QY 712 GGGTGTGCTCTGACATGCCAGGGGTCTTGTTGCTGTGCTG-TGTTCCAGAAAGCTT 770
DB 41248 GGGTGTGCTCTAGGAATGCCAGGGGTCTGTAACAGATGGGTTTTCCAGAGAGGTTT 41307
QY 771 CTGCACTACAGATGCTTTGGGAGGAGTAAAGCATATGCTGATTTTCCGTTGCCCTGG 830
DB 41308 CTGCACTATGGGTCTGTGTGAGGGGAGGCAATCTGCTGATTTTCTGTGCTCTGG 41367
QY 831 CTAGTACCCCTACAGAGAAATAGTGGGTAGGAGGAG-GGGCGAGCGGCTGCTGAC 889
```

Db 41368 CTAGTACCCCCCTTACGAGAAAGATTAAGCGCTAAGCCAGAGGGGGGAGACGCCCACTACAC 41427

Qy 890 ATCTGCGCTGCTCTTATCACTTATCATATAAGGAGAAAGTATGATTGCGATAC 949

Db 41428 ATGCTGGCTGCTCTTATCACTTATCATATAAGGAGAAAGTATGATTGCGATAC 41487

Qy 950 TGACACTGTATGACTGAGGAGAAACAAAGAGCTCCGTGACAGCTCTGAGTATGACAGG 1009

Db 41488 TGACACTGTATGACTGAGGAGAAACAAAGAGCTCCGTGAGTATGATAAGC 41547

Qy 1010 CTGCTCTCTCTCTCTGACGACG 1032

Db 41548 TGACACAGCCCTCCCAATGACG 41570

RESULT 5  
AC091095/c  
LOCUS  
DEFINITION Homo sapiens chromosome 8 clone RP11-77H21 map 8, WORKING DRAFT  
AC091095  
SEQUENCE, 22 unordered pieces.  
AC091095  
VERSION AC091095.2 GI:14626897  
HTG; HTGS PHASE1; HTGS\_DRAFT; HTGS\_FULLTOP.  
KEYWORDS  
SOURCE Homo sapiens (human)  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Primates; Catherhini;  
Homnidae; Homo.  
REFERENCE  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N., Anderson, S.,  
Barn, N., Bastien, V., Boguslavsky, L., Bouckhalter, B., Brown, A.,  
Cammer, J., Chang, J., Choepel, Y., Colangelo, M.,  
Collins, S., Collymore, A., Cooke, P., Darelano, K., Dewar, K.,  
Diaz, J., Dodge, S., Fato, S., Ferreira, P., Fitzhugh, W., Gage, D.,  
Galgan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L.,  
Grand-Pierre, N., Hagos, B., Heatford, A., Horton, L., Hulme, W.,  
Iliev, I., Johnson, R., Jones, C., Karatas, A., LaRoque, K.,  
Lamazares, R., Landers, T., Lehotzky, J., Levine, R., Liu, G.,  
Maclean, C., MacDonald, P., Marguis, N., Matthews, C., McCarthy, M.,  
McEwan, P., McKernan, K., McPheters, R., Meldrum, J., Meneus, L.,  
Mihova, T., Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Norbu, C.,  
Norman, C.H., O'Connor, T., O'Donnell, P., O'Neill, D., Oliver, J.,  
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,  
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,  
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupbach, R., Seaman, S.,  
Severy, P., Sougniez, C., Spencer, B., Stange-Thomann, N.,  
Stojanovic, N., Straus, N., Subramanian, A., Talamas, J., Tesfaye, S.,  
Theodore, J., Travers, M., Travis, N., Triggillo, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J., Young, G.,  
Zaitoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (29-MAR-2001) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Jul 8, 2001 this sequence version replaced gi:13487963.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE  
JOURNAL  
COMMENT

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence.submissions@genome.wi.mit.edu  
Project Information  
Center project name: 77\_H\_21  
Center clone name: L10790  
Summary Statistics  
Sequencing vector: Plasmid; n/a; 100% of reads  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.960731

Consensus quality: 174312 bases at least Q40  
Consensus quality: 178687 bases at least Q30  
Consensus quality: 180102 bases at least Q20  
Insert size: 163000; agarose-fp  
Insert size: 181115; sum-of-contigs  
Quality coverage: 9.2 in Q20 bases; agarose-fp  
Quality coverage: 8.3 in Q20 bases; sum-of-contigs

NOTE: This is a 'working draft' sequence. It currently  
consists of 22 contigs. The true order of the pieces  
is not known and their order in this sequence record is  
arbitrary. Gaps between the contigs are represented as  
runs of N, but the exact sizes of the gaps are unknown.  
This record will be updated with the finished sequence  
as soon as it is available and the accession number will  
be preserved.

1 49245: contig of 49245 bp in length  
49245: gap of 100 bp  
49346: contig of 611 bp in length  
49346: gap of 100 bp  
49957: contig of 683 bp in length  
50057: gap of 100 bp  
50739: contig of 683 bp in length  
50740: gap of 100 bp  
50839: contig of 719 bp in length  
51558: gap of 100 bp  
51559: contig of 915 bp in length  
51659: gap of 100 bp  
52573: contig of 915 bp in length  
52574: gap of 100 bp  
52673: contig of 588 bp in length  
53261: gap of 100 bp  
53262: contig of 654 bp in length  
53361: gap of 100 bp  
54015: contig of 654 bp in length  
54016: gap of 100 bp  
54116: contig of 653 bp in length  
54768: gap of 100 bp  
54769: contig of 1193 bp in length  
54869: gap of 100 bp  
56061: contig of 111 bp in length  
56162: gap of 100 bp  
57272: contig of 725 bp in length  
57373: gap of 100 bp  
58097: contig of 725 bp in length  
58098: gap of 100 bp  
58197: contig of 801 bp in length  
58198: gap of 100 bp  
58998: contig of 167 bp in length  
59099: gap of 100 bp  
60815: contig of 1440 bp in length  
60816: gap of 100 bp  
62255: contig of 1440 bp in length  
62256: gap of 100 bp  
62355: contig of 994 bp in length  
63349: gap of 100 bp  
63350: contig of 787 bp in length  
64236: gap of 100 bp  
64237: contig of 1250 bp in length  
64337: gap of 100 bp  
65626: contig of 243 bp in length  
65726: gap of 100 bp  
68149: contig of 243 bp in length  
68150: gap of 100 bp  
78644: contig of 1095 bp in length  
78645: gap of 100 bp  
78744: contig of 1399 bp in length  
78745: gap of 100 bp  
94443: contig of 1399 bp in length  
94444: gap of 100 bp  
94543: contig of 61042 bp in length  
94544: gap of 100 bp  
155685: contig of 27530 bp in length  
155686: gap of 100 bp  
183215: contig of 27530 bp in length.  
Location/Qualifiers  
1.183215  
/organism="Homo sapiens"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9606"  
/chromosome="8"  
/map="8"  
/clone="RP11-77H21"  
/clone\_11b="RP11-77H21 Human Male BAC"  
1.49245  
/note="assembly\_fragment  
clone end:SP6  
vector\_side:left"  
49246.49345

## FEATURES

misc\_feature  
gap

misc_feature	/estimated_length=100 49346. .49956 /note="assembly_fragment"
gap	49957. .50056 /estimated_length=100
misc_feature	50057. .50739 /note="assembly_fragment"
gap	50740. .50839 /estimated_length=100
misc_feature	50840. .51558 /note="assembly_fragment"
gap	51559. .51658 /estimated_length=100
misc_feature	51659. .52573 /note="assembly_fragment"
gap	52574. .52673 /estimated_length=100
misc_feature	52674. .53261 /note="assembly_fragment"
gap	53262. .53361 /estimated_length=100
misc_feature	53362. .54015 /note="assembly_fragment"
gap	54016. .54115 /estimated_length=100
misc_feature	54116. .54768 /note="assembly_fragment"
gap	54769. .54868 /estimated_length=100
misc_feature	54869. .56061 /note="assembly_fragment"
gap	56062. .56161 /estimated_length=100
misc_feature	56162. .57272 /note="assembly_fragment"
gap	57273. .57372 /estimated_length=100
misc_feature	57373. .58097 /note="assembly_fragment"
gap	58098. .58197 /estimated_length=100
misc_feature	58198. .58998 /note="assembly_fragment"
gap	58999. .59098 /estimated_length=100
misc_feature	59099. .60715 /note="assembly_fragment"
gap	60716. .60815 /estimated_length=100
misc_feature	60816. .62255 /note="assembly_fragment"
gap	62256. .62355 /estimated_length=100
misc_feature	62356. .63349 /note="assembly_fragment"
gap	63350. .63449 /estimated_length=100
misc_feature	63450. .64236 /note="assembly_fragment"
gap	64237. .64336 /estimated_length=100
misc_feature	64337. .65626 /note="assembly_fragment"
gap	65627. .65726 /estimated_length=100
misc_feature	65727. .68149 /note="assembly_fragment"
gap	68150. .68249 /estimated_length=100
misc_feature	68250. .78644 /note="assembly_fragment"

19.6%; Score 239; DB 14; Length 183215;

	Best Local Similarity	65.1%	Pred. No. 1.7e-44;	
	Matches 537;	Conservative 0;	Mismatches 250;	Indels 38; Gaps 11;
QY	244	AAGGAAAAAGGTGATGTTGTTGTTCCCTCGACATACCTCACTCACTCAATATCTATGTTACT	303	
Db	118402	AAAAAAGTTGTAATTCCTCTTCTCTACAGACAGCTTAAATTTCAATTTCTCTATTTGTA	118344	
QY	304	TAAACAGCATTTATATCGGCGCAACCACTTAAAGTTTATTTATTTATATGTAATTTTGTTG	363	
Db	118342	TGTATATGACTTATATCATGATGATGTAATTTTATGAGTAATTCAGAGTGAAGTTT	118283	
QY	364	AGGCACAGCATGTGAGTGTCTGAGAAACATGATGTGTACTGATTTATATA-----	417	
Db	118282	AAGA-A-ATAAAGTCAGAGCTCAG-AACTGATGTTTCTTACCTGACTTAATTAACAGATC	118222	
QY	418	----TATCAACCTGGGTTAAATTAATA-----AAATGTGCATTAACCTTAAAAAAA	466	
Db	118224	TTTCTATCCAACTGTTTCGAAATGAATAATCTTTTGTATACCTGTGCTGATCAAGATC	118165	
QY	467	ACCAATATACCAACAGACTTATCTACCTCCCTGGAAAAAGCATTTATCAAGG---CTGAT	523	
Db	118164	AATTAATATCATGAATCTATTTTTCCTCTGTAAGAAAGCACAGCTTTACAGAGCCAAAC	118105	
QY	524	CTTACGCTTTATATTTTACATTAAGAAAAATAAACCAAGTCCGATATAGCTGTAATTT	583	
Db	118104	TTCTTCTCTCTTATTTGTATATAAATAAATAAATTAACCAAGTCCGATCTTATTTGTATATT	118045	
QY	584	ATTCCTA-----AAGAACAAGAACTTCTACATAGCTTTAAATTAAGTATTTACT	636	
Db	118044	TTTATTTCTTAAGAAAGAAAAACAGAACTTTCATTTGACTTCAACATTTAAAGTTATTA	637	
QY	637	CAGATATCTGCAAGCTTACCTTACCAACAGAGAG--ACAGACAACAGAGCCCA-GCTA	693	
Db	117984	CAGATATTTTTCAGAGCTTACAGACGCAAAAAATCAGTTTACAGCAAAAGATCACTGCT	117925	
QY	694	CTCTCTAGAAA---TAATTAGGTTGTTGCTCTGACATGCCAGGGTCTTTGAGCTG	749	
Db	117924	CTCTCTAGAAAATACTTAATTTGGGGTGTGCTTAGGAAATGCCAGGGGCTCTGTAAACAG	117865	
QY	750	GTCCTGAG-TTCCAGAAAGGCTTCTGCAATACACAGTCTTTGGGGAGTAAGACATATGC	808	
Db	117864	ATCGGTTTTTCCACAGAGGTTTCTGACAGATGGGTTCTGTTGGAAGGACGAGATTTCTC	117805	
QY	809	TCTGATTTTCTGTGTTGCTGCTAGTGAACCCCTTACAGAAAGATATAGGTGAGCCAG	868	
Db	117804	TCTGATTTTCTGTGTTGCTGCTAGTGAACCCCTTACAGAAAGATATAGGTGAGCCAG	117745	
QY	869	-GGGGGAGGCGGCTGAGCATATCTGGGCTCTTATCAACTTATATATAAGGAA	927	
Db	117744	AGGGCGAGACAGCCCATACATATGTGGGCTCTTATCAACTTATATATAAGGAA	117685	
QY	928	GGAAAGTATGATTCGATACCTGACACTGTATGATCTCAGGGGAGAGAAACAAAGATCCGTG	987	
Db	117684	GGAAAGTATGATTCGATACCTGACACTGTATGATCTGAGGAGAGAGAAACAAAGATCCGTG	117625	
QY	988	CAGACCTCTGAGATGACAGAGGCTGCTCTTCTCTCAGAGAGC	1032	
Db	117624	GTGAAGGCTGCTGTAAAGCTGACACAGCCCTCCCAAGTGAAGC	117580	
RESULT 6				
AX087869/c				
LOCUS	AX087869	240825 bp	DNA	linear
DEFINITION	Sequence 1 from Patent WO0114550.			
ACCESSION	AX087869			
VERSION	AX087869.1	GI:13396862		
KEYWORDS	.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;			
	Homidae; Homo.			
REFERENCE	1			



```
AUTHORS      Barry,C., Chumakov,I. and Blumenfeld,M.
TITLE         Prostate cancer-related gene 3 (pg3) and diallelic markers thereof
JOURNAL       Patent: WO 0114550-A 1 01-MAR-2001;
              GENSER (FR)
FEATURES
  source      Location/Qualifiers
              1. .240825
               /organism="Homo sapiens"
               /mol_type="unassigned DNA"
               /db_xref="taxon:9606"
  misc_feature 1. .2000
               /note="5,regulatory region"
  primer_bind  1823. .1840
               /note="5-390.pu"
  primer_bind  1980. .1998
               /note="5-390-177.mis"
  misc_binding 1987. .2011
               /bound_moiety="5-390-177.probe"
  variation    1999
               /note="5-390-177 : polymorphic base G or C"
  primer_bind  2000. .2018
               /note="5-390-177.mis complement"
  exon         2001. .2079
               /note="exon A"
  primer_bind  2108. .2125
               /note="5-390.rp complement"
  primer_bind  4559. .4577
               /note="5-391.pu"
  primer_bind  4582. .4600
               /note="5-391-43.mis"
  misc_binding 4589. .4613
               /bound_moiety="5-391-43.probe"
  variation    4601
               /note="5-391-43 : polymorphic base A or G"
  primer_bind  4602. .4620
               /note="5-391-43.mis complement"
  exon         4627. .4718
               /note="exon B"
  primer_bind  4891. .4908
               /note="5-391.rp complement"
  primer_bind  10007. .10025
               /note="5-392.pu"
  exon         10115. .10233
               /note="exon C"
  primer_bind  10209. .10227
               /note="5-392-222.mis"
  misc_binding 10216. .10240
               /bound_moiety="5-392-222.probe"
  variation    10228
               /note="5-392-222 : polymorphic base G or T"
  primer_bind  10229. .10247
               /note="5-392-222.mis complement"
  primer_bind  10267. .10285
               /note="5-392-280.mis"
  misc_binding 10274. .10298
               /bound_moiety="5-392-280.probe"
  variation    10286
               /note="5-392-280 : polymorphic base G or T"
  primer_bind  10287. .10305
               /note="5-392-280.mis complement"
  variation    10370
               /note="5-392-364 : insertion of G"
  primer_bind  10411. .10430
               /note="5-392.rp complement"
  exon         26810. .26897
               /note="exon D"
  exon         31357. .31471
               /note="exon E"
  exon         34261. .34404
               /note="exon F"
  exon         37377. .37466
               /note="exon S"
  primer_bind  39556. .39574
               /note="4-59.rp"
  exon         39704. .40858
               /note="exon T"
  primer_bind  39877. .39896
               /note="4-58.rp"
  primer_bind  39925. .39943
               /note="4-58-318.mis"
  misc_binding 39932. .39956
               /bound_moiety="4-58-318.probe"
  variation    39944
               /note="4-58-318 : polymorphic base G or T"
  primer_bind  39945. .39963
               /note="4-58-318.mis complement"
  primer_bind  39953. .39970
               /note="4-59.pu complement"
  primer_bind  39954. .39972
               /note="4-58-289.mis"
  misc_binding 39961. .39985
               /bound_moiety="4-58-289.probe"
  variation    39973
               /note="4-58-289 : polymorphic base G or C"
  primer_bind  39974. .39992
               /note="4-58-289.mis complement"
  primer_bind  40242. .40259
               /note="4-58.rp complement"
  primer_bind  41137. .41154
               /note="4-54.rp"
  primer_bind  41366. .41384
               /note="4-54-199.mis"
  misc_binding 41373. .41397
               /bound_moiety="4-54-199.probe"
  primer_bind  41385. .41403
               /note="4-54-180.mis"
  variation    41385
               /note="4-54-199 : polymorphic base A or C"
  primer_bind  41386. .41404
               /note="4-54-199.mis complement"
  misc_binding 41392. .41416
               /bound_moiety="4-54-180.probe"
  variation    41404
               /note="4-54-180 : polymorphic base A or C"
  primer_bind  41405. .41423
               /note="4-54-180.mis complement"
  primer_bind  41564. .41581
               /note="4-54.pu complement"
  primer_bind  42122. .42141
               /note="4-51.rp"
  primer_bind  42213. .42231
               /note="4-51-312.mis"
  misc_binding 42220. .42244
               /bound_moiety="4-51-312.probe"
  variation    42232
               /note="4-51-312 : polymorphic base G or C"
  primer_bind  42233. .42251
               /note="4-51-312.mis complement"
  primer_bind  42526. .42543
               /note="4-51.pu complement"
  exon         50436. .50545
               /note="exon G"
  primer_bind  67289. .67309
               /note="99-86.rp"
  primer_bind  67456. .67474
               /note="99-86-266.mis"
  misc_binding 67463. .67487
               /bound_moiety="99-86-266.probe"
  variation    67475
               /note="99-86-266 : polymorphic base A or C"
  primer_bind  67476. .67494
               /note="99-86-266.mis complement"
  primer_bind  67724. .67741
               /note="99-86.pu complement"
  primer_bind  69182. .69200
               /note="4-88.rp"
  primer_bind  69502. .69520
```

misc_binding	/note="4-88-107.mis" 69509. .69533 /bound_moiety="4-88-107.probe"
variation	69521 /note="4-88-107 : polymorphic base A or G" 69522. .69540 /note="4-88-107.mis complement"
primer_bind	69609. .69626 /note="4-88.pu complement"
primer_bind	72698. .72715 /note="5-397.pu"
primer_bind	72819. .72837 /note="5-397-141.mis"
misc_binding	72826. .72850 /bound_moiety="5-397-141.probe"
variation	72838 /note="5-397-141 : polymorphic base G or T" 72839. .72857 /note="5-397-141.mis complement"
exon	72881. .72918 /note="exon H"
primer_bind	73099. .73117 /note="5-397.tp complement"
primer_bind	75858. .75877 /note="5-398.pu"
exon	75989. .76151 /note="exon I"
primer_bind	76041. .76059 /note="5-398-203.mis"
misc_binding	76048. .76072 /bound_moiety="5-398-203.probe"
variation	76060 /note="5-398-203 : polymorphic base A or C"
primer_bind	76061. .76079 /note="5-398-203.mis complement"
primer_bind	76289. .76306 /note="5-398.tp complement"
primer_bind	81006. .81025 /note="99-12738.pu"
primer_bind	81234. .81252 /note="99-12738-248.mis"
misc_binding	81241. .81265 /bound_moiety="99-12738-248.probe"

Query	Match	Similarity	19.5%	Score 238	DB 6	Length 240825
Best Local	Similarity	64.3%	Pred. No. 2.9e-44			
Match	545	Conservative	0	Mismatches 265	Indels 38	Gaps 11
QY	221	CCCGAGAAACCCGTATACAGTGAAGAAAGNGTATGTTGGTCTCTCGACATAC	280			
Db	158572	CTCAAAAAAAAAAAAAAAAAAAAAAAAAAGTTGTATCTCTTTTTCACACGACCTT	158513			
QY	281	TCACTCAAATATCTATGTGTTACTTAAACAGCAATTAATCAGGCCAACACTTAAAGTTT	340			
Db	158512	AAATTTCTAATTTCTATTGTATGTATGTAATGAATCACTTATCAGTACAGTGATTTTAAAGT	158453			
QY	341	TATTTTATATGATTTTGTGTTAAGCACAGACATGTGAGTCTGTGAAAACTGATGTTG	400			
Db	158452	GAATATCCAGNGTGAAGTTTTTTTAAAGA-ATAAAGTCACAGCTCAG-AAACGTAGTTTT	158395			
QY	401	GTAACCTGATTTTATTA-----TATCAACTGGGTTTAAATTAATA-----AA	443			
Db	158394	CTAACGTGACTTATTAACAGATCTTCTCTATCCAACTGTTGGAAATGAAATCTTTGTG	158335			
QY	444	AATGTCATTACTTAAAAAAAAAACCAATACACACAGACTTTACTCCCTTGAAAA	503			
Db	158334	ACTTGCGTACGTGCTCAAGATCAACTTAATCAAGAAATCTATTTTCCCTCTGAAAG	158275			
QY	504	GCACATTTTCAAGGG---CTGATCTTAAGCTTTATATTTTTCATTAAGAAATTAACCA	560			
Db	158274	GCACACGTTTACAGAGCCAAACTCTTCTCTTTTATTTGTATTAACAAAAATTAACCGA	158215			
QY	561	GGTCCCGATATAGCTGTAAATTTTATTCCTA-----AAAGAACAGAACTTCACTATG	613			

Db	158214	AGTCCGACCTTATTTGTAGTATTTTATTCTTAAAGAAAAACAGAACTTTCATTGTA	1581555
Qy	614	CTTTAAATTAAGTANTTACCTCAGATACCTGACAGCTTACCCACAAACGACAG--	671
Db	158154	CTTCAACATTAAAGTATTATTAACCTCAGATATTTTGGCAGCTTAGACAGGCAAAATCAGTT	158095
Qy	672	ACAGACAAACAGAGCCCCA-GCTACTCTCTAGAGAAA---TAAATAGGGGTGGCTCTGA	726
Db	158094	TCAGACAAAAGATCACTACTGCTCTCTTAGAGAAAATCTTAATTTGGGGTGGTCCTAGGA	158035
Qy	727	CATGCCACAGGGGCTCTGTGGCTGTCTG-TGTTCCAGAGGCTTTCGACATACACAGTC	785
Db	158034	AATGCCACAGGGGTCCTGTAAACAGATCGTATTTTCCAGAGGGTTCTGCACATGGGTCC	157975
Qy	786	CTTTGGGGCAGTAAGACATATGCTCTGATTTTTCCTGTGTGCTGCTAGTGAACCCCTAC	845
Db	157974	TGGTTGGAGGGCAGGACATTTCTCTCTGATTTTTCCTGTGCTGCTAGTGAACCCCTAC	157915
Qy	846	AGGAAGATAGTGGGTGAGCCAGG-GGGCGGACGGGTGGCTGCACATGTCTGGCTCTC	904
Db	157914	AGGAAGATACCGCTTAAGCCAGAGAGGGCGAGACCCACTACATGTCTGGCTGCTCT	157855
Qy	905	TATCACTTATATATTAAGGAGAGAAAGTATGATTTGATCTGATCAGACTGTAGACTC	964
Db	157854	TATCAACTTATATATTAAGGAGAGAAAGTATGATTTGATCTGATCAGACTGTAGAGATC	157795
Qy	965	AGGGAGAAACAAAGATCCGTCGACAGCCTCTGGAGTAGACAGGGGTGCTCTCTCTC	1024
Db	157794	TGGGAGAGAGAGAACAAAGAGACCGTGAAAGCTGTCTGTAAAGCTGACACAGCCTCC	157735
Qy	1025	AGGACAGC 1032	
Db	157734	AAGTGAGC 157727	
RESULT 7			
AX523960/c	AX523960	240825 bp	linear
LOCUS	AX523960	240825 bp	linear
DEFINITION	Sequence 1 from Patent WO0206641.		
ACCESSION	AX523960		
VERSION	AX523960.1	GI:25168889	
KEYWORDS			
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Barry,C. and Chumakov,I.		
AUTHORS	Pg-3 and diallelic markers thereof		
TITLE	Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;		
JOURNAL	Homidae; Homo.		
GENSET (FR)			
FEATURES			
source	Location/Qualifiers		
	1..240825		
	/organism="Homo sapiens"		
	/mol_type="unassigned DNA"		
	/db_xref="taxon:9606"		
	1..2000		
misc_feature	/note="5'regulatory region"		
	1823..1840		
primer_bind	/note="5-390.pu"		
	1980..1998		
primer_bind	/note="5-390-177.mis"		
	1987..2011		
misc_binding	/bound_molecly="5-390-177.probe"		
	1999		
variation	/note="5-390-177 : polymorphic base G or C"		
	2000..2018		
primer_bind	/note="5-390-177.mis complement"		
	2001..2079		
exon	/note="exon A"		
	2108..2125		

```
primer_bind /note="5-390.rp complement"
4559.4577
/note="5-391.pu"
primer_bind 4582.4600
/note="5-391-43.mis"
misc_binding 4589.4613
/bound_moiety="5-391-43.probe"
variation 4601
/note="5-391-43 : polymorphic base A or G"
primer_bind 4602.4620
/note="5-391-43.mis complement"
4627.4718
/note="exon B"
primer_bind 4891.4908
/note="5-391.rp complement"
10007.10025
/note="5-392.pu"
exon 10115.10233
/note="exon C"
primer_bind 10209.10227
/note="5-392-222.mis"
10216.10240
/bound_moiety="5-392-222.probe"
variation 10228
/note="5-392-222 : polymorphic base G or T"
10229.10247
/note="5-392-222.mis complement"
primer_bind 10267.10285
/note="5-392-280.mis"
10274.10298
/bound_moiety="5-392-280.probe"
variation 10286
/note="5-392-280 : polymorphic base G or T"
10287.10305
/note="5-392-280.mis complement"
10370
/note="5-392-364 : insertion of G"
10411.10430
/note="5-392.rp complement"
26810.26897
/note="exon D"
exon 31357.31471
/note="exon E"
34261.34404
/note="exon F"
37377.37466
/note="exon S"
39556.39574
/note="4-59.rp"
39704.40858
/note="exon T"
39877.39896
/note="4-58.rp"
39925.39943
/note="4-58-318.mis"
39932.39956
/bound_moiety="4-58-318.probe"
variation 39944
/note="4-58-318 : polymorphic base G or T"
39945.39963
/note="4-58-318.mis complement"
39953.39970
/note="4-59.pu complement"
39954.39972
/note="4-58-289.mis"
39961.39985
/bound_moiety="4-58-289.probe"
variation 39973
/note="4-58-289 : polymorphic base G or C"
39974.39992
/note="4-58-289.mis complement"
40242.40259
/note="4-58.pu complement"

primer_bind 41137.41154
/note="4-54.rp"
41366.41384
/note="4-54-199.mis"
misc_binding 41373.41397
/bound_moiety="4-54-199.probe"
primer_bind 41385.41403
/note="4-54-180.mis"
41385
/note="4-54-199 : polymorphic base A or C"
41386.41404
/note="4-54-199.mis complement"
41392.41416
/bound_moiety="4-54-180.probe"
variation 41404
/note="4-54-180 : polymorphic base A or C"
41405.41423
/note="4-54-180.mis complement"
41564.41581
/note="4-54.pu complement"
42122.42141
/note="4-51.rp"
42213.42231
/note="4-51-312.mis"
42220.42244
/bound_moiety="4-51-312.probe"
variation 42232
/note="4-51-312 : polymorphic base G or C"
42233.42251
/note="4-51-312.mis complement"
42526.42543
/note="4-51.pu complement"
50436.50545
/note="exon G"
67289.67309
/note="99-86.rp"
67456.67474
/note="99-86-266.mis"
67463.67487
/bound_moiety="99-86-266.probe"
67475
/note="99-86-266 : polymorphic base A or G"
67476.67494
/note="99-86-266.mis complement"
67724.67741
/note="99-86.pu complement"
69182.69200
/note="4-88.rp"
69502.69520
/note="4-88-107.mis"
69509.69533
/bound_moiety="4-88-107.probe"
69521
/note="4-88-107 : polymorphic base A or G"
69522.69540
/note="4-88-107.mis complement"
69609.69626
/note="4-88.pu complement"
72698.72715
/note="5-397.pu"
72819.72837
/note="5-397-141.mis"
72826.72850
/bound_moiety="5-397-141.probe"
72838
/note="5-397-141 : polymorphic base G or T"
72839.72857
/note="5-397-141.mis complement"
72861.72918
/note="exon H"
73099.73117
/note="5-397.rp complement"
75858.75877
```

```

exon /note="5-398.pu"
75989..76151
primer_bind /note="exon 1"
76041..76059
misc_binding /note="5-398-203.mls"
76048..76072
/bound_moiety="5-398-203.probe"
variation /note="5-398-203 : polymorphic base A or C"
76060
primer_bind /note="5-398-203.mls complement"
76061..76079
primer_bind /note="5-398-203.mls complement"
76289..76306
primer_bind /note="5-398.tp complement"
81006..81025
primer_bind /note="99-12738.pu"
81234..81252
primer_bind /note="99-12738-248.mls"
81241..81265
misc_binding /bound_moiety="99-12738-248.probe"

Query Match 19.5%; Score 238; DB 6; Length 240825;
Best Local Similarity 64.3%; Pred. No. 2.8e-44;
Matches 545; Conservative 0; Mismatches 265; Indels 38; Gaps 11;

QY 221 CCCGAGAAACCCCTGATACAGTGAAGAAAGGTGATCTGTTCCCTCTGACATACT 280
DB 158572 CTCAAAAAAGAAAAAAGAAAAAAGTTGATTCCTTTCTACACAGACACTT 158513

QY 281 TCACTCAATATCTATTTGTTACTTAAACAGACATTTATCAGGCCAAACACTTAACTT 340
DB 158512 AAATCTTAATTTCTCTATTTGTAATGTAAGAACTTATATGACATGATTTTAAAGT 158453

QY 341 TATTTGTATGATTTTGTGTTAAGCAGACAGATGTGAGTCTGAGAAACTGATGTTG 400
DB 158452 GAATATCCAAAGTGAGATTTTAAAGA-ATTAAGGTGAGACTGAG-AACTGATGTTT 158395

QY 401 GTAATCTGATTTAATTA-----TATCAACCTGGGTTAAATAA-----AA 443
DB 158394 CTAAAGTGCTTAATTAACAGATCTTTCTATCAACCTGTCGAAATGAATAATCTTTTGT 158335

QY 444 AATGTGCAATTAATTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 503
DB 158334 ACTGGCTATGTTGTTCAAGATATCAATAAATAAATAAATAAATAAATAAATAA 158275

QY 504 GCACATTTACAAAGG--CTGATCTTAGCCTTTAATTTACATTAAGAAATAAACC 560
DB 158274 GCACAGTTTACAGGAGCCAACTCTCTCTTTATTTGTAATAAATAAATAAACC 158215

QY 561 GGTCCCGATATAGCTGATTTTATTTCCTA-----AAAGAACAATACTTCACTATG 613
DB 158214 AGTCCGACCTATTTGTAATTTTATTTCTTAAGAAATAAATAAATAAATAAATAA 158155

QY 614 CTTTAAATTAAGTATTAATCTCAGATATCTGCAAGCTTACCAAGAGAGAG-- 671
DB 158154 CTTCAACATTAAGTATTAATCTCAGATATTTTGCAGCTTACAGCGAAATAAGTT 158095

QY 672 ACAAGACAAGAGCCCCA-GCTACTCTAGAGAA---TAATAGGGTGGTCTCTGA 726
DB 158094 TCAGACAAAAAGATCACTGCTCTCTAGAGAAATACTTAATAGGGGTTGCTTACAG 158035

QY 727 CATGCCAAGGGGCTTGGGCTGGTCTG--TGTTCCAGAAAGGCTTGTAGTACAGTAC 785
DB 158034 AATGCCAAGGGGCTTGTACAGATGCTTTTCCAGAGGGGTTTGTGACATGGGCTCC 157975

QY 786 CTTTGGGGCAGTAAGACATATGCTGATTTTCTGTTGCTGCTAGTGAACCCCTTAC 845
DB 157974 TGGTTGAGAGGAGCAGCATTTCTGCTGATTTTTCGTTGCTGCTAGTGAACCCCTAC 157915

QY 846 AGGAAGATATGTGGGTAGCCAGG--GGCGGAGCGGCTGGCTGACATGTTCTGGCTCTCT 904
DB 157914 AGGAAGATATGCGCTAAGCGCAGAGGGCGGAGCGCCACTACATGTTCTGGCTCTCTCT 157855

QY 905 TATCAACTATATATTAAGGAAGAAAGATGATTGATTGGGATATCGACATCTGTACATC 964

```

```

DB 157854 TATCAACTTATCATTAATAAGAAAGAAAGTATGATTCGATACCTGACATGAGATC 157795
QY 965 AGGGAGAAACAAAGAGTCCCTGAGACCTCTGAGTGAAGAGGAGGCTGCTCTCTC 1024
DB 157794 TGGGAGAGAGAAACAAAGAACCGTGAAGCTGCTCTGTAAGCTGACAGCCCTCC 157735
QY 1025 AGGACAGC 1032
DB 157734 AAGTGAGC 157727

RESULT 8
AY563557 6714 bp DNA linear PRI 19-AUG-2004
LOCUS Homo sapiens angiotensin-2 (AGP2) gene, promoter region, 5' UTR
DEFINITION AY563557
ACCESSION AY563557
VERSION AY563557.1 GI:50082537
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
Hominae; Homo.
REFERENCE 1 (bases 1 to 6714)
Hegen,A., Koide,S., Weindel,K., Marne,D., Augustin,H.G. and
Fiedler,U.
Expression of Angiotensin-2 in Endothelial Cells is Controlled by
Positive and Negative Regulatory Promoter Elements
(ser) Arterioscler. Thromb. Vasc. Biol. (2004) In press
15284088
JOURNAL PUBMED 2 (bases 1 to 6714)
Fiedler,U., Hegen,A., Augustin,H.G. and Koide,S.
Direct Submission
Submitted (04-MAR-2004) Vascular Biology and Angiogenesis Research,
Tumor Biology Center Freiburg, Breisacher Str. 117, Freiburg 79106,
Germany
FEATURES
source Location/Qualifiers
1..6714
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="8"
/map="BP23.1"
1..>6714
/gene="AGP2"
1..4427
/gene="AGP2"
promoter /gene="AGP2"
4428..>5192
mRNA /gene="AGP2"
/product="angiotensin-2"
4428..4904
/gene="AGP2"
4905..>5192
CDS /gene="AGP2"
/note="Tie2-ligand"
/codon_start=1
/product="angiotensin-2"
/protein_id="AAT69979.1"
/db_xref="GI:50082538"
/translation="WMQIVFPLSGDVLAAAYNNFRKSMDSIGKQYOVHSGSYT
FLPEMDNCRSSSPYVSNVORDAPLEIDSVQRLOVLENNMTQMLMK"

ORIGIN
Query Match 19.5%; Score 237.4; DB 8; Length 6714;
Best Local Similarity 65.0%; Pred. No. 4.8e-44;
Matches 536; Conservative 0; Mismatches 251; Indels 38; Gaps 11;

QY 244 AAGGAAGAGTATGTTGTTTCTCTGACATATCTCACTCAATATCTATTTACT 303
DB 3856 AAAAAAGTTGATTCCTCTTTCTACACAGACACTTAATTTCTATTTGTA 3915

```





Kowis, C., Kraft, C.L., Lebow, H., Levay, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorenschew, L., Loulseged, H., Lozano, R., Lu, X., Ma, J., Maheshwari, M., Mahindratne, M., Mahmoud, M., Malloy, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Mawhinney, S., McLeod, M.P., McNeill, T.Z., Meenan, E., Mitoavajevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Munidasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaelelemeh, O., Okwuonu, G., Olarnpungsoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L., Puoz, M., Quiroz, J., Rachlin, E., Reeves, K., Regier, M.A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S., Sanders, W., Saverly, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C.D., Smales, D., Sneed, A., Sodergren, E., Song, X.-Z., Sorelle, R., Soes, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabors, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejos, Z., Umanat, K., Valae, R., Vera, V., Villaseana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willson, R., Wleczek, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, I., Yoon, V., Yu, R., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D.R., Holt, R.A., Smith, H.O., Weinstock, G., and Gibbs, R.A.

Unpublished  
2 (bases 1 to 221218)  
Direct Submision  
Submitted (16-NOV-2004) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

3 (bases 1 to 221218)  
Cow Genome Sequencing Consortium.  
Direct Submision  
Submitted (01-JUL-2005) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA

On Jun 26, 2005 this sequence version replaced gi:58037895. The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/rac/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: hgsc-help@bcm.tmc.edu  
----- Project Information  
Center project name: FANL  
Center clone name: CH240-8B21  
----- Summary Statistics  
Assembly program: Atlas 3.0;  
Consensus quality: 213012 bases at least Q40  
Consensus quality: 21532 bases at least Q30  
Consensus quality: 21744 bases at least Q20  
Estimated insert size: 21423; sum-of-contigs estimation  
Quality coverage: 7x in Q20 bases; sum-of-contigs estimation  
-----  
\* NOTE: Estimated insert size may differ from sequence length  
\* (see [http://www.hgsc.bcm.tmc.edu/docs/genbank\\_draft\\_data.html](http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html)).  
\* NOTE: This is a 'working draft' sequence. It currently

consists of 27 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1  
2088: contig of 2088 bp in length  
2089 2188: gap of 50 bp  
2139 11896: contig of 9758 bp in length  
11897 11946: gap of 50 bp  
11947 13770: contig of 1824 bp in length  
13771 13820: gap of 50 bp  
13821 17181: contig of 3361 bp in length  
17182 17231: gap of 50 bp  
17232 18621: contig of 1390 bp in length  
18622 18671: gap of 50 bp  
18672 61815: contig of 43144 bp in length  
61816 61865: gap of 50 bp  
61866 65984: contig of 4119 bp in length  
65985 66034: gap of 50 bp  
66035 87544: contig of 21510 bp in length  
87545 87594: gap of 50 bp  
87595 95737: contig of 8143 bp in length  
95738 95787: gap of 50 bp  
95788 100716: contig of 4929 bp in length  
100717 100766: gap of 50 bp  
100767 108628: contig of 7862 bp in length  
108629 108678: gap of 50 bp  
108679 109836: contig of 1158 bp in length  
109837 109886: gap of 50 bp  
109887 119073: contig of 9187 bp in length  
119074 119123: gap of 50 bp  
119124 127103: contig of 7980 bp in length  
127104 127153: gap of 50 bp  
127154 134419: contig of 7266 bp in length  
134420 134469: gap of 50 bp  
134470 155171: contig of 20702 bp in length  
155172 155221: gap of 50 bp  
155222 190589: contig of 35368 bp in length  
190590 190639: gap of 50 bp  
190640 201726: contig of 11087 bp in length  
201727 201826: gap of unknown length  
201827 203687: contig of 1861 bp in length  
203688 203737: gap of 50 bp  
203738 211192: contig of 7455 bp in length  
211193 212346: gap of unknown length  
212347 212446: contig of 1054 bp in length  
212447 213466: gap of unknown length  
213467 213566: contig of 1020 bp in length  
213567 214871: gap of unknown length  
214872 214971: contig of 1305 bp in length  
214972 214972: gap of unknown length  
214973 216198: contig of 1227 bp in length  
216199 216298: gap of unknown length  
216299 218080: contig of 1782 bp in length  
218081 218180: gap of unknown length  
218181 220028: contig of 1848 bp in length  
220029 220128: gap of unknown length  
220129 221218: contig of 1090 bp in length.  
Location/Qualifiers  
1..221218  
/organism="Bos taurus"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:9913"  
/clone="CH240-8B21"  
2089..2138  
/estimated\_length=50  
11897..11946  
/estimated\_length=50  
13771..13820  
/estimated\_length=50  
17182..17231  
gap

```

gap      /estimated_length=50
          18622. .18671
          /estimated_length=50
gap      61816. .61865
          /estimated_length=50
gap      65985. .66034
          /estimated_length=50
gap      87545. .87594
          /estimated_length=50
gap      95738. .95787
          /estimated_length=50
gap      100717. .100766
          /estimated_length=50
gap      108629. .108678
          /estimated_length=50
gap      109837. .109886
          /estimated_length=50
gap      119074. .119123
          /estimated_length=50
gap      127104. .127153

```

Query Match 15.6%; Score 190.6; DB 14; Length 221218;  
 Best Local Similarity 67.6%; Pred. No. 2.4e-33;  
 Matches 390; Conservative 0; Mismatches 169; Indels 18; Gaps 8;

```

QY 432 AAAAAAATGCTAATAAAAAAAACCAATACCAAGACTTACTT 491
DB 141965 AGTTTGAAGTACTGCTTAAAGTTCAAAATCAAGTAAATCAAGCAAACTATTTC 141906
QY 492 CCCCTTGAA--AAGCATTACAAAGGCTGATCTTAAGCTTTATTTTCAATTAAG 548
DB 141905 CCAATGAAAGGAGAGAGTTTCAAGAGGAGACTTCTTCTTTGTTGTAATTAACA 141846
QY 549 AAAAAATCAAGAGTCCGATATA--GCTTAATTTTATTCCTAAGAGCAAGAACTT 605
DB 141845 AAAATTAACCAAGTCTCAATATTGCAATTTTATTTCTTAAGGAAA--AAACAG 141788
QY 606 TCATCTGCTTAAATTAAGGATTAACCTGAGTACTGCAAGCTTACCTTACAAAC 665
DB 141787 GAACTTCAATGATCTTAAAGTATTGCCCCAGATATTACAGCTTGGCGCTGTGAAA 141728
QY 666 GAGCAGACAGACAAAGAGCCCGAGCTACTCTCT--AGGAATATTAATTAAGGAGTGTG 719
DB 141727 TCAGTTTCAGACAGAGAGACCTACAGGCTCTCTCAGAGAAACAGTATTAGGGTGTG 141668
QY 720 CCTTGAACATGCCAGGGCTTTGTGGCTGTGTG--TTCCAGAAAGCTTCTGCAAGTA 778
DB 141667 CCTTGAATTAACCCAGGGGTCTGTAACTGATCGGTTTTTCCAGAGGGCTTCTGACAGA 141608
QY 779 CACAGT-CCTTTGGGGCAGTAAGCACTATGCTGTATTTTCTGTGCTGGCTAGTGA 837
DB 141607 TATAGTCCCGGTGAGCAGTACATCAAGCCCGGATTTTCTGTGGCTGGTGTGA 141548
QY 838 -CCCCCTACAGAAAGTATGAGTGAAGCCAGG-GGGCGAGCGGCTGCTGACATGTCT 895
DB 141547 CCCCCCTACAGAAAGTATGAGTGAAGCCAGGCGGAGGGCGGAGCGGCTGACATGTCT 141488
QY 896 GGCTGCTCTTATCACTTATCATATTAAGGAGAAAGTATGATTGGATCTGACAC 955
DB 141487 GCTGCTCTTATCACTTATCATATTAAGGAGAAAGTATGATTGGATCTGACAC 141428
QY 956 TGTAGACTGAGGAGAAACAAAGATCCGTGACAGC 992
DB 141427 TGCAGATCTGGGGAAGAGACAAAGCAGCGCTTAAC 141391

```

RESULT 12  
 HSI64G8F  
 LOCUS HSI64G8F 290 bp DNA linear PRI 18-OCT-1995  
 DEFINITION H.sapiens Cpg island DNA genomic MseI fragment, clone 164g8,  
 forward read cpg164g8.f1a.  
 ACCESSION 257119  
 VERSION 257119.1 GI:1028350

# KEYWORDS

Cpg island; genomic MseI fragment.

# SOURCE

Homo sapiens (human)

# ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;  
 Homnidae; Homo.

# REFERENCE

1 Cross,S.H., Charlton,J.A., Nan,X. and Bird,A.P.  
 Purification of Cpg islands using a methylated DNA binding column  
 Nat. Genet. 6 (3), 236-244 (1994)

# JOURNAL

8012384

# PUBMED

2 (bases 1 to 290)

# AUTHORS

Dodsworth,S.J., Huckle,E., Wilkinson,P. and Micklem,G.

# TITLE

Submitted (16-OCT-1995) The Sanger Centre, Hinxton, Cambridgeshire,

# JOURNAL

CB10 1RQ, England. E-mail contact: humquery@sanger.ac.uk

# COMMENT

Vector: pGBM-5Zf(-)

# FEATURES

Clones are available from the UK MRC Human Genome Mapping Project  
 Resource Centre, Hinxton, Cambridgeshire CB10 1RQ, UK. See URL:  
 http://www.hgmp.mrc.ac.uk/ for details  
 or contact: biohelp@hgmp.mrc.ac.uk.

# source

```

1..290
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/clone="164g8"
/sex="male"
/lisase_type="blood"
/clone_lib="CGI-1"
/dev_stage="adult"

```

# ORIGIN

Query Match 14.4%; Score 176.2; DB 8; Length 290;  
 Best Local Similarity 81.9%; Pred. No. 7.5e-30;  
 Matches 227; Conservative 0; Mismatches 48; Indels 2; Gaps 2;

```

QY 706 TAATTAGGAGTGTGCTTGAATGCCCCAGAGGCTTGTGGCTGTGTG--TTCCAGAGA 764
DB 1 TAATTAGGAGTGTGCTTGAATGCCCCAGAGGCTTGTGGCTGTGTG--TTCCAGAGA 60
QY 765 AGGCTTGCAGTACACAGTCTTTGGGGCAGTAAAGCACTATGCTGATTTTCCGTGT 824
DB 61 GGGTTTTCAGACATGGGCTCTGTGTGAGGGCAGGATTCGTCTGATTTTTCGTGT 120
QY 825 GCTGTGCTAGTACCCCTTACAGAGAGATAGTGTGAGCCAGG-GGGCGAGCGGCTGG 883
DB 121 GCTGTGCTAGTACCCCTTACAGAGAGATAGTGTGAGCCAGG-GGGCGAGCGGCTGG 180
QY 884 CTGACATGTCTGGCTGCTTATCAACTTATCATATTAAGGAGAAAGTATGATTTC 943
DB 181 CTACACATGTCTGGCTGCTTATCAACTTATCATATTAAGGAGAAAGTATGATTTC 240
QY 944 GGATACAGACCTGATAGCTCAGGGGAGAAACAAAGA 980
DB 241 GGATACAGACCTGATAGCTCAGGGGAGAGAAACA 277

```

RESULT 13  
 AC141003  
 LOCUS AC141003 132392 bp DNA linear HTG 27-MAR-2003  
 DEFINITION Rattus norvegicus clone CH230-526115, \*\*\* SEQUENCING IN PROGRESS  
 \*\*\*, 64 unordered pieces.  
 AC141003  
 AC141003.1 GI:28875862  
 VERSION HTG; HTGS PHASE1.  
 KEYWORDS Rattus norvegicus (Norway rat)  
 SOURCE Rattus norvegicus  
 ORGANISM Rattus norvegicus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
 Sciurognathi; Muridae; Muridae; Murinae; Rattus.  
 1 (bases 1 to 132392)  
 Muzny,D.,Marie., Metzker,M.,Lee., Abramson,S., Adams,C., Alder,J.,



Allen, C., Allen, H., Alsbrooks, S., Amin, A., Anguiano, D., Anyalebechi, V., Aoyagi, A., Ayodeji, M., Baca, E., Baden, H., Baldwin, D., Bandaranaike, D., Barber, M., Barnstead, M., Benhamed, F., Biswal, K., Blair, J., Blankenburg, K., Blyth, P., Brown, M., Bryant, N., Buhay, C., Burch, P., Burrell, K., Calderon, E., Cardenas, V., Carter, K., Cavazos, I., Caesar, H., Center, A., Chacko, J., Chavez, D., Chen, G., Chen, R., Chen, Y., Chen, Z., Chu, J., Cleaveland, C., Cockrell, R., Cox, C., Coyle, M., Cree, A., D'Souza, L., Davila, M. L., Davis, S., Davy-Carroll, L., De Anda, C., Dederich, D., Delgado, A., Dennis, C., Deramo, C., Ding, Y., Dinh, H., Ditya, K., Draper, H., Dugan-Rocha, S., Dunn, A., Durbin, K., Duval, B., Eaves, K., Egan, A., Escotto, M., Eugene, C., Evans, C. A., Falls, T., Fan, G., Fernandez, S., Finley, M., Flagg, N., Forbes, L., Foster, M., Foster, P., Frazer, C. M., Gabisi, A., Ganta, R., Garcia, M., Garner, T., Garza, M., Gebregorgis, E., Geer, K., Gill, R., Girdy, M., Guerra, M., Guevara, W., Gunaratne, P., Haaland, W., Hamil, C., Hamilton, C., Hamilton, J., Harvey, Y., Havlak, P., Hawes, A., Henderson, N., Hernandez, J., Hernandez, R., Hines, S., Hladun, S. L., Hodgson, A., Hognes, M., Hollins, B., Howells, S., Hulyk, S., Hume, J., Idelbird, D., Jackson, A., Jackson, L., Jacob, L., Jiang, H., Johnson, B., Johnson, R., Jolivet, A., Karpathy, S., Kelly, S., Kelly, S., Khan, Z., King, L., Kovar, C., Kowals, C., Kraft, C. L., Lebow, H., Levan, J., Lewis, L., Li, Z., Liu, J., Liu, J., Liu, W., Liu, Y., London, P., Longacre, S., Lopez, J., Lorensbueh, L., Louisedge, H., Lozado, R. J., Lu, X., Ma, J., Maheshwari, M., Maindarcine, M., Mahmoud, M., Mallory, K., Mangum, A., Mangum, B., Mapua, P., Martin, K., Martin, R., Martinez, E., Manning, S., McLeod, M., McNeill, T., Meenen, E., Milosavljevic, A., Miner, G., Minja, E., Montemayor, J., Moore, S., Morgan, M., Morris, K., Morris, S., Mundaasa, M., Murphy, M., Nair, L., Nankervis, C., Neal, D., Newton, N., Nguyen, N., Norris, S., Nwaokelenen, O., Okunolu, G., Olarunpungoon, A., Pal, S., Parks, K., Pasternak, S., Paul, H., Perez, A., Perez, L., Pfankoch, C., Plopper, F., Poindexter, A., Popovic, D., Primus, E., Pu, L.-L., Puzo, M., Quiroz, T., Rachlin, E., Reeves, K., Regier, M. A., Reigh, R., Reilly, B., Reilly, M., Ren, Y., Reuter, M., Richards, S., Riggs, F., Rives, C., Rodkey, T., Rojas, A., Rose, M., Rose, R., Ruiz, S. J., Sanders, W., Savery, G., Scherer, S., Scott, G., Shatsman, S., Shen, H., Shetty, J., Shvartsbeyn, A., Sisson, I., Sitter, C. D., Smajs, D., Snead, A., Sodergren, E., Song, X.-Z., Stortelle, R., Soes, J., Steinle, M., Strong, R., Sutton, A., Svatek, A., Tabor, P., Taylor, C., Taylor, T., Thomas, N., Thomas, S., Tingey, A., Trejoe, Z., Uemami, K., Valae, R., Vera, V., Villaana, D., Waldron, L., Walker, B., Wang, J., Wang, Q., Wang, S., Warren, J., Warren, R., Wei, X., White, F., Williams, G., Willison, R., Wleczyk, R., Wooden, H., Worley, K., Wright, D., Wright, R., Wu, J., Yakub, S., Yen, J., Yoon, D., Yoon, V., Yu, F., Zhang, J., Zhou, J., Zhou, X., Zhao, S., Dunn, D., von Niederhausen, A., Weiss, R., Smith, D. R., Holt, R. A., Smith, H. O., Weinstein, G., and Gibbs, R. A.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Unpublished  
2 (bases 1 to 132392)  
Worley, K.C.  
Direct Submission  
Submitted (07-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
(bases 1 to 132392)  
Worley, K.C.  
Direct Submission  
Submitted (27-MAR-2003) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA  
-----  
Genome Center  
Center: Baylor College of Medicine  
Center code: BCM  
Web site: <http://www.hgsc.bcm.tmc.edu/>  
Contact: [hgsc-help@bcm.tmc.edu](mailto:hgsc-help@bcm.tmc.edu)  
-----  
Project Information  
Center project name: GXKS  
Center clone name: CH230-526115  
-----  
Summary Statistics  
Sequencing vector: Plasmid  
Chemistry: Dye-terminator Big Dye; 100% of reads  
Assembly program: Phrap; version 0.990329

Consensus quality: 98530 bases at least Q40	
Consensus quality: 105678 bases at least Q30	
Consensus quality: 110765 bases at least Q20	
Estimated insert size: 86014; sum-of-contigs estimation	
Quality coverage: 1x in Q20 bases; sum-of-contigs estimation	
NOTE: Estimated insert size may differ from sequence length (see <a href="http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html">http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html</a> ).	
NOTE: This is a 'working draft' sequence. It currently consists of 64 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.	
1	1018: contig of 1018 bp in length
1019	1118: gap of unknown length
1119	2164: contig of 1046 bp in length
2165	2264: gap of unknown length
2265	3340: contig of 1076 bp in length
3341	3440: gap of unknown length
3441	4445: contig of 1005 bp in length
4446	4545: gap of unknown length
4546	5558: contig of 1013 bp in length
5559	5658: gap of unknown length
5659	6671: contig of 1013 bp in length
6672	6771: gap of unknown length
6772	7849: contig of 1078 bp in length
7850	7949: gap of unknown length
7950	9069: contig of 1120 bp in length
9070	9169: gap of unknown length
9170	10301: contig of 1132 bp in length
10302	10401: gap of unknown length
10402	11937: contig of 1536 bp in length
11938	12037: gap of unknown length
12038	13406: contig of 1369 bp in length
13407	13506: gap of unknown length
13507	14990: contig of 1484 bp in length
14991	15090: gap of unknown length
15091	16107: contig of 1017 bp in length
16108	16207: gap of unknown length
16208	17285: contig of 1078 bp in length
17286	17385: gap of unknown length
17386	18963: contig of 1578 bp in length
18964	19063: gap of unknown length
19064	20794: contig of 1731 bp in length
20795	20894: gap of unknown length
20895	22416: contig of 1522 bp in length
22417	22516: gap of unknown length
22517	23639: contig of 1123 bp in length
23640	23739: gap of unknown length
23740	24842: contig of 1103 bp in length
24843	24942: gap of unknown length
24943	26028: contig of 1086 bp in length
26029	26128: gap of unknown length
26129	27718: contig of 1590 bp in length
27719	27818: gap of unknown length
27819	29221: contig of 1403 bp in length
29222	29321: gap of unknown length
29322	30448: contig of 1127 bp in length
30449	30548: gap of unknown length
30549	31894: contig of 1346 bp in length
31895	31994: gap of unknown length
31995	33378: contig of 1384 bp in length
33379	33478: gap of unknown length
33479	34875: contig of 1397 bp in length
34876	34975: gap of unknown length
34976	36002: contig of 1027 bp in length
36003	36102: gap of unknown length
36103	37499: contig of 1397 bp in length
37500	37599: gap of unknown length
37600	39671: contig of 2072 bp in length



Best Local Similarity 62.1%; Pred. No. 1.1e-21;  
Matches 295; Conservative 0; Mismatches 164; Indels 16; Gaps 4;

```
QY 522 ATCTTAGCCTTATATTTACAATAAAGAAATAAACCAAGTCCCGATATAGCTGTAAT 581
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1530 ATTTTATTTTATTTATTTATTAATAAATAAATCGAAGTTTGATTTATTTGTATATA 1589
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 582 TTATTCCT-----AAAGAACAGAACTTTCACATAGCTTAAATTAAGTGATAC 634
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1590 TTTTATTTTAAAGGAAAAAATAGAAATTTTATTTATTTATTAATTAAGTTATAT 1649
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 635 CTCAGTACTCTGCAAGCTTAGCCTACAAACGACAGACAGACAAACAGACCCCACTAC 694
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1650 TTTAGATATTTTGTAGTTAGTACGTAATAAATTAGTTTACATTAAGAGATTATTTG 1709
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 695 TCTCT-----AGGAAATATATTAGGTTGCTGCTGACATGCCAGGGGCTTGTGGC 747
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1710 TTTTATTTAGAAATATTTAATTGGGTGTGTGTAGAAATGTTTAGGGGTTTGTATAT 1769
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 748 TGGTCTG-TGTTCCCAAGAGGCTTCTGCAGTACAGTCCCTTGGGCGATAGCACTAT 806
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1770 AGATCGTATTTTATAGAGGTTTTGTAGTATGGGTTTTGTTGAGGGTATGTTT 1829
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 807 GCTCTGATTTTCTGCTGCTAGTGAACCCCTACAGGAAATAGGGTGAGCCA 866
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1830 GTTTTGATTTTGTGTGTGTGTAGTATTTTATAGGAGATTAACGGTTAAGTTA 1889
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 867 GG-GGGCGAGCGGCTGCTGCACATGCTGCTGCTTATCACTTATCATATTAAGG 925
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1890 GGAAGGCGAGTAGTTATTAATATATGTTTGTGTTTATTAATTAATTAAGGA 1949
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 926 AAGGAAAGTATGATTGATTCGATACAGACTGTAAGCTCAGGGAGAAACAAGA 980
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 1950 AAGGAAAGTATGATTCGATATGATATTGTAGGATTTGGGAGAGAGAAATA 2004
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

Search completed: December 12, 2005, 14:03:08  
Job time : 6377 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 12, 2005, 10:36:32 ; Search time 757 Seconds  
(without alignments)  
10740.984 Million cell updates/sec

Title: US-10-627-075-1

Perfect score: 1220  
Sequence: 1 ggcgtctccagctaccgacgc.....ccaaaaagagagagaaatcg 1220

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4996997 segs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

N\_Geneseq\_21:\*  
1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003cs:\*  
11: geneseqn2003ds:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*  
14: geneseqn2005s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1220	100.0	1220	12	AD043367 Mouse Ang
2	276	22.5	2475	14	AD61847 Murine Ag
3	238	19.5	240823	10	AD669391 Human PG-
4	238	19.5	240825	4	AAFP24497 Aaf24497 Human PG-
5	238	19.5	240825	6	ABQ81802 Abq81802 Human PG-
6	213	17.5	2424	12	AD009378 Ad009378 Murine An
7	213	17.5	2424	13	AD513802 Ad513802 Mouse ang
8	213	17.5	2424	13	AD339008 Ad339008 Mouse Ang
9	211.8	17.4	62705	12	AD009371 Ad009371 Human Ang
10	156.6	12.8	3251	8	ABX63024 Abx63024 Human CDN
11	143.4	11.8	763	10	AD69421 Ad69421 Human Ang
12	140.6	11.5	2846	6	AB067157 Ab067157 Human ang
13	139.4	11.4	2846	6	AB067158 Ab067158 Human ang
14	58	4.8	2000	8	AB471938 Ab471938 Rice gene
15	55	4.5	10710	6	AB132893 Ab132893 Human imm
16	54.4	4.5	15698	6	AB134141 Ab134141 Human imm
17	52.6	4.3	2265	14	ADW10467 Adw10467 Colon pro
18	51.6	4.2	6523	6	ABN80121 Abn80121 Human che
19	51.6	4.2	6523	10	AD684216 Ad684216 Human lym

C 20	51.6	4.2	6523	10	AD684140 Ad684140 Human lym
C 21	51.6	4.2	6523	13	AD689620 Ad689620 Oligonuc1
C 22	51.6	4.2	6523	13	AD689346 Ad689346 Oligonuc1
C 23	51.6	4.2	6823	6	AB133144 Ab133144 Human imm
C 24	51.2	4.2	20486	6	AB134611 Ab134611 Human met
C 25	51.2	4.2	20486	7	AD699872 Ad699872 Complemen
C 26	50.8	4.2	6101	13	AD689690 Ad689690 Oligonuc1
C 27	50.8	4.2	6101	13	AD689416 Ad689416 Oligonuc1
C 28	50.2	4.1	9964	6	AB132099 Ab132099 Human imm
C 29	50	4.1	4415	6	AB133906 Ab133906 Human met
C 30	50	4.1	4415	6	AB134602 Ab134602 Human met
C 31	50	4.1	4415	6	AB170423 Ab170423 Chemical1
C 32	50	4.1	4415	6	AD699863 Ad699863 Bisulphit
C 33	49.8	4.1	15479	6	ABK39964 AbK39964 Human che
C 34	49.2	4.0	13784	6	ABK40062 AbK40062 Human che
C 35	49	4.0	2000	11	AC137108 AC137108 Rice stre
C 36	49	4.0	6127	6	AB134449 Ab134449 Human met
C 37	49	4.0	6127	6	AB170120 Ab170120 Chemical1
C 38	49	4.0	6127	7	AD699710 Ad699710 Complemen
C 39	49	4.0	8758	6	AB133118 Ab133118 Human imm
C 40	49	4.0	8900	13	AD689685 Ad689685 Oligonuc1
C 41	49	4.0	8900	13	AD689411 Ad689411 Oligonuc1
C 42	48.6	4.0	487	9	ACH18710 Ach18710 Human adu
C 43	48.4	4.0	4661	6	ABK31226 Abk31226 Signal tr
C 44	48.4	4.0	4661	6	AB170185 Ab170185 Chemical1
C 45	48.4	4.0	4661	6	AA61136 Aa61136 Human gen

#### ALIGNMENTS

RESULT 1	
AD043367	AD043367 standard; DNA; 1220 BP.
XX	XX
AC	AD043367;
XX	XX
DT	29-JUL-2004 (first entry)
DE	Mouse Ang-2 cis-acting regulatory sequence.
XX	XX
KW	Ang-2; enhancer; mouse; cell imaging; antiangiogenic; ds.
XX	XX
OS	Mus sp.
XX	XX
FH	Key Location/Qualifiers
FT	prim_transcript 1008..1220
FT	/*tag= a
FT	/partial
FT	/note= "Includes translation initiation codon at 3' end"
XX	XX
PN	W02004037975-A2.
PD	PD
XX	XX
PD	06-MAY-2004.
XX	XX
XX	25-JUL-2003; 2003WO-US023152.
XX	XX
PR	25-JUL-2003; 2002US-0398583P.
PR	24-JUL-2002; 2003US-00627075.
XX	XX
PA	(DAND ) DANA FARBER CANCER INST INC.
XX	XX
PI	Livington DM, Kung AL;
XX	XX
DR	WPI; 2004-365502/34.
XX	XX
PT	New polynucleotide comprising a first nucleic acid encoding a light-
PT	generating gene product and a second nucleic acid encoding a selectable
PT	marker, useful in identifying a compound capable of modulating
PT	angiogenesis.
XX	XX
PS	Disclosure; SEQ ID NO 1; 69pp; English.
XX	XX

CC The present sequence is that of the cis-acting regulatory region of the  
 CC murine Ang-2 gene located 5' to the start of transcription. This cis-  
 CC acting DNA regulates preferential expression in endothelial muscle cells  
 CC of a polypeptide-encoding DNA to which it is operably linked. The  
 CC invention provides compositions that include a polynucleotide encoding a  
 CC reporter gene (e.g., a light-generating moiety), a polynucleotide encoding a  
 CC selectable marker (e.g., an antibiotic) and optionally a regulatory  
 CC element (e.g., the Ang-2 cis-acting regulatory sequence, or a portion of  
 CC it that retains the ability to effect transcription of operably linked  
 CC sequences in endothelial cells). The composition is used for imaging  
 CC cells. The cells may be healthy cells or tumor cells, in which case  
 CC tumour growth, engraftment and/or metastasis can be monitored by  
 CC observing light emitted from the light-generating gene product. Also  
 CC provided are vectors comprising the polynucleotides, cells (e.g., primary  
 CC cells, cells from established cell lines, or tumor cells) comprising the  
 CC vector or polynucleotides, and transgenic animals and methods for their  
 CC production. The transgenic animals are used in methods for the  
 CC identification of a compound capable of modulating angiogenesis or which  
 CC has tumour inhibitor activity, or which is a modulator of a signal  
 CC transduction pathway.

XX Sequence 1220 BP; 370 A; 254 C; 295 G; 301 T; 0 U; 0 Other;

Query Match 100.0%; Score 1220; DB 12; Length 1220;

Best Local Similarity 100.0%; Pred. No. 1e-312;

Matches 1220; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAGCTTTCCAGACGATCTTCGACATTAATCTTACTGATGAAAGGGGTGATCTC 60  
 Db 1 GAGCTTTCCAGACGATCTTCGACATTAATCTTACTGATGAAAGGGGTGATCTC 60  
 QY 61 TGGACGACGAGACGACGAGAGTGTGTTTGAAGTCACTGAGGCCCAAGTTT 120  
 Db 61 TGGACGACGAGACGACGAGAGTGTGTTTGAAGTCACTGAGGCCCAAGTTT 120  
 QY 121 ATGGCCAGGGGCTTTGAACTTAATTAAGGGGAAAGTATTGGCTGAGCCACTGAC 180  
 Db 121 ATGGCCAGGGGCTTTGAACTTAATTAAGGGGAAAGTATTGGCTGAGCCACTGAC 180  
 QY 181 TGGGACTAATTTAATGAGAAATGCGACAGAGTGTGAGCCGAGAAACCTGTATACA 240  
 Db 181 TGGGACTAATTTAATGAGAAATGCGACAGAGTGTGAGCCGAGAAACCTGTATACA 240  
 QY 241 GTGAGGAAAAGGTGATGTTTCTCTGACATACCTGCAATATCTATTGTT 300  
 Db 241 GTGAGGAAAAGGTGATGTTTCTCTGACATACCTGCAATATCTATTGTT 300  
 QY 301 ACTTAACAGACAATTAATGAGGCAACCACTTAAGTTTATTTGATATGATTTG 360  
 Db 301 ACTTAACAGACAATTAATGAGGCAACCACTTAAGTTTATTTGATATGATTTG 360  
 QY 361 TTAAGGCACAGCATGTGAGTCTGAGAAACTGATGTTGGTATCTTATTAATAT 420  
 Db 361 TTAAGGCACAGCATGTGAGTCTGAGAAACTGATGTTGGTATCTTATTAATAT 420  
 QY 421 CAAACTGGGTTAAATAAAAAATGTCATTAATCTTAATAAATAACCAATACCA 480  
 Db 421 CAAACTGGGTTAAATAAAAAATGTCATTAATCTTAATAAATAACCAATACCA 480  
 QY 481 AGACTTTAATCTCCCTTGGAAAAAGCAATTTAAGGGCTGATCTTATCTTATTTA 540  
 Db 481 AGACTTTAATCTCCCTTGGAAAAAGCAATTTAAGGGCTGATCTTATCTTATTTA 540  
 QY 541 CAATTAAGAAAATAAACCAAGTCCGATATAGCTGATATTTATCTTAAGAAGACA 600  
 Db 541 CAATTAAGAAAATAAACCAAGTCCGATATAGCTGATATTTATCTTAAGAAGACA 600  
 QY 601 AACTTCTACTATGCTTAAATAATTAAGTATTAATCTGATATCTGCAAGCTTAGCCTA 660  
 Db 601 AACTTCTACTATGCTTAAATAATTAAGTATTAATCTGATATCTGCAAGCTTAGCCTA 660  
 QY 661 CAACGACGACGACAGACAGAGCCGAGCTACTCTAGAGAAATATTTAGGGTGGTGC 720  
 Db 661 CAACGACGACGACAGACAGAGCCGAGCTACTCTAGAGAAATATTTAGGGTGGTGC 720

Db 661 CAACGACGACGACAGACAGAGCCGAGCTACTCTAGAGAAATATTTAGGGTGGTGC 720  
 QY 721 CTCTGACATGCCAGAGGGCTTTGTGGCTGTGTGTGTCCAGAAAGCTTCTGACATACA 780  
 Db 721 CTCTGACATGCCAGAGGGCTTTGTGGCTGTGTGTGTCCAGAAAGCTTCTGACATACA 780  
 QY 781 CAGTCTTTGGGGCAGTAAAGCACTATAGCTTGATTTTCCGTGTCCTGGCTAGTACCC 840  
 Db 781 CAGTCTTTGGGGCAGTAAAGCACTATAGCTTGATTTTCCGTGTCCTGGCTAGTACCC 840  
 QY 841 CCTACAGGAATAGTGGTGTAGCCAGGGGGCGAGCGCTGTGTCACATGTCTGGCTG 900  
 Db 841 CCTACAGGAATAGTGGTGTAGCCAGGGGGCGAGCGCTGTGTCACATGTCTGGCTG 900  
 QY 901 CTCTTATCACTTATCATATTAAGGAAGAAAGTATTTGATTCGGATCTGACACTGTAG 960  
 Db 901 CTCTTATCACTTATCATATTAAGGAAGAAAGTATTTGATTCGGATCTGACACTGTAG 960  
 QY 961 ACTCAGGGGAGAAACMAAGAGTCCGTGACAGCTCTGAGTGTGACAGAGGCTGCTCTTCC 1020  
 Db 961 ACTCAGGGGAGAAACMAAGAGTCCGTGACAGCTCTGAGTGTGACAGAGGCTGCTCTTCC 1020  
 QY 1021 TCTCAGACAGCTCCGATGTGTCCGGGGAGAGAGAGAGACAGAGCACTGGGAAA 1080  
 Db 1021 TCTCAGACAGCTCCGATGTGTCCGGGGAGAGAGAGAGAGACAGAGCACTGGGAAA 1080  
 QY 1081 GAGCCTGTGGGGACGAGAGAGGCTCTCACTGATGACTTATTCACAGGCAAGCCCT 1140  
 Db 1081 GAGCCTGTGGGGACGAGAGAGGCTCTCACTGATGACTTATTCACAGGCAAGCCCT 1140  
 QY 1141 GTGCTTTGACAGCAGCTGAGAGCTCAGAGCGCAAGTTTGCTGAATCTCAGAGTTTAAAC 1200  
 Db 1141 GTGCTTTGACAGCAGCTGAGAGCTCAGAGCGCAAGTTTGCTGAATCTCAGAGTTTAAAC 1200  
 QY 1201 CCAAAAAGAGAGAGAGATG 1220  
 Db 1201 CCAAAAAGAGAGAGAGATG 1220

RESULT 2  
 AD261847  
 ID AD261847 standard; DNA, 2475 BP.  
 XX  
 AC AD261847;  
 XX  
 DT 30-JUN-2005 (first entry)  
 XX  
 DE Murine Asp2 gene, SEQ ID 1571.  
 XX  
 KW Drug screening; Peroxisome Proliferator-Activated Receptor gamma;  
 KW PPAR gamma; gene; ds.  
 XX  
 OS Mus musculus.  
 XX  
 PN US2005084872-A1.  
 PD 21-Apr-2005.  
 XX  
 PF 23-JAN-2004; 2004US-00764420.  
 XX  
 PR 24-JAN-2003; 2003US-0442797P.  
 PR 30-MAY-2003; 2003US-0474413P.  
 XX  
 PA (LUMP/) LUM P Y.  
 PA (TANY/) TAN Y.  
 PA (DAIH/) DAI H.  
 PA (MUIS/) MUISE E S.  
 PA (BERG/) BERGER J P.  
 PA (THOM/) THOMPSON J R.  
 XX  
 PI Lum PY, Tan Y, Dai H, Muise ES, Berger JP, Thompson JR;  
 XX WPI; 2005-313963/32.  
 DR



```
Db 158032 AATGCCAGGGGCTCTGTAACAGATCGGTTTTTCCCAAGAGGTTCTGCACATGGGATCC 157973
Qy 786 CTTTGGGGCAGTAAGCACTATGCTGTGATTTTTCCTGTGCTGAGTAAGACCCCTAC 845
Db 157972 TGGTTGAGGGGAGGCACTTCGCTGTGATTTTTCCTGTGCTGAGTAAGACCCCTAC 157913
Qy 846 AGGAAGATAGTGGGTGACCCAGG-GGGCGAGCCGGCTGGCTGCAATGTCTGCTGCTCT 904
Db 157912 AGGAAGATTAACGGCTTAAGCCAGAGGGCGGAGCAGCCACTACATGTCTGGCTGCTCT 157853
Qy 905 TATCAACTTATCATATTAAGGAAGAAAGTATTCGATTCGATCTGACACTGTAGACTC 964
Db 157852 TATCACTTATCATATTAAGGAAGAAAGTATTCGATTCGATCTGACACTGTAGGATC 157793
Qy 965 AGGGAGAAACAAAGAGTCCGTCGACACTCTGAGTAGAGAGGCTGCTCCTTCCTCTC 1024
Db 157792 TGGGGAGAGAGAACAAAGACCGTGAAGCTGCTGTAAAGCTGACACAGCCCTCCC 157733
Qy 1025 AGGACAGC 1032
Db 157732 AAGTGAGC 157725
```

```
RESULT 4
AAF24497/c
ID AAF24497 standard; cDNA; 240825 BP.
XX
XX AAF24497;
XX
XX 23-MAY-2001 (first entry)
XX
XX Human PG-3 gene.
XX
XX Human PG-3 gene.
XX
XX Human; PG-3; cancer; BRCA1; chromosome 8p23; ds.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
XX FT misc_feature 1..2000
XX FT /*tag= a
XX FT /note= "5' regulatory region"
XX FT primer_bind 1823..1840
XX FT /*tag= b
XX FT primer_bind 1980..1998
XX FT /*tag= c
XX FT misc_binding 1987..2011
XX FT /*tag= d
XX FT /note= "binds probe"
XX FT allele replace(1999,C)
XX FT /*tag= e
XX FT primer_bind complement(2000..2018)
XX FT /*tag= f
XX FT CDS 2001..238825
XX FT /*tag= g
XX FT /product= "PG-3"
XX FT /note= "this sequence contains introns"
XX FT exon 2001..2079
XX FT /*tag= h
XX FT primer_bind /label= A
XX FT primer_bind 2108..2125
XX FT /*tag= i
XX FT primer_bind 4559..4577
XX FT /*tag= j
XX FT primer_bind 4582..4600
XX FT /*tag= k
XX FT misc_binding 4589..4613
XX FT /*tag= l
XX FT /note= "binds probe"
XX FT allele replace(4601,G)
XX FT /*tag= m
XX FT primer_bind complement(4602..4620)
XX FT /*tag= n
```

```
FT exon 4627..4718
FT /*tag= o
FT /label= B
FT primer_bind 4891..4908
FT /*tag= p
FT primer_bind 10007..10025
FT /*tag= q
FT exon 10115..10233
FT /*tag= r
FT primer_bind 10209..10227
FT /*tag= s
FT misc_binding 10216..10240
FT /*tag= t
FT allele /note= "binds probe"
FT replace(10228,T)
FT /*tag= u
FT primer_bind complement(10229..10247)
FT /*tag= v
FT primer_bind 10267..10285
FT /*tag= w
FT misc_binding 10274..10298
FT /*tag= x
FT /note= "binds probe"
FT replace(10286,T)
FT /*tag= y
FT complement(10287..10305)
FT /*tag= z
FT allele replace(10370,n)
FT /*tag= aa
FT primer_bind 10411..10430
FT /*tag= ab
FT exon 26810..26897
FT /*tag= ac
FT /label= D
FT exon 31357..31471
FT /*tag= ad
FT /label= E
FT exon 34261..34404
FT /*tag= ae
FT /label= F
FT exon 37377..37466
FT /*tag= af
FT /label= S
FT primer_bind 39556..39574
FT /*tag= ah
FT exon 39704..40858
FT /*tag= ag
FT /label= T
FT primer_bind 39877..39896
FT /*tag= ai
FT primer_bind 39925..39943
FT /*tag= aj
FT misc_binding 39932..39956
FT /*tag= ak
FT /note= "binds probe"
FT allele replace(39944,T)
FT /*tag= al
FT primer_bind complement(39945..39963)
FT /*tag= am
FT primer_bind 39953..39970
FT /*tag= an
FT primer_bind 39954..39972
FT /*tag= ao
FT misc_binding 39961..39985
FT /*tag= ap
FT /note= "binds probe"
FT allele replace(39973,C)
FT /*tag= aq
FT primer_bind complement(39974..39992)
FT /*tag= ar
FT primer_bind 40242..40259
FT /*tag= as
```



```
FT primer_bind 41137..41154
FT /*tag= ac
FT primer_bind 41366..41384
FT /*tag= au
FT misc_binding 41373..41397
FT /*tag= av
FT /note= "binds probe"
FT primer_bind 41385..41403
FT /*tag= ax
FT allele replace(41385,C)
FT /*tag= aw
FT primer_bind complement(41386..41404)
FT /*tag= ay
FT misc_binding 41392..41416
FT /*tag= az
FT /note= "binds probe"
FT allele replace(41404,C)
FT /*tag= ba
FT primer_bind complement(41405..41423)
FT /*tag= bd
FT primer_bind 41564..41581
FT /*tag= bc
FT primer_bind 42122..42141
FT /*tag= bd
FT primer_bind 42213..42231
FT /*tag= be
FT misc_binding 42220..42244
FT /*tag= bf
FT /note= "binds probe"
FT allele replace(42232,C)
FT /*tag= bg
FT primer_bind complement(42233..42251)
FT /*tag= bh
FT primer_bind 42526..42543
FT /*tag= bi
FT exon 50436..50545
FT /*tag= bj
FT /label= G
FT primer_bind 67289..67309
FT /*tag= bk
FT primer_bind 67456..67474
FT /*tag= bl
FT misc_binding 67463..67487
FT /*tag= bm
FT /note= "binds probe"
FT allele replace(67475,G)
FT /*tag= bn
FT primer_bind complement(67476..67494)
FT /*tag= bo
FT primer_bind 67724..67741
FT /*tag= bp
FT primer_bind 69182..69200
FT /*tag= bq
FT primer_bind 69502..69520
FT /*tag= br
FT misc_binding 69509..69533
FT /*tag= bs
FT /note= "binds probe"
FT allele replace(69521,G)
FT /*tag= bt
FT primer_bind complement(69522..69540)
FT /*tag= bu
FT primer_bind 69609..69626
FT /*tag= bv
FT primer_bind 72698..72715
FT /*tag= bw
FT primer_bind 72819..72837
FT /*tag= bx
FT misc_binding 72826..72850
FT /*tag= by
FT /note= "binds probe"
FT allele replace(72838,T)
FT /*tag= bz
```

```
FT primer_bind complement(72839..72857)
FT /*tag= ca
FT exon 72881..72918
FT /*tag= cb
FT /label= H
FT primer_bind 73099..73117
FT /*tag= cc
```

Query Match 19.5%; Score 238; DB 4; Length 240825;

Best Local Similarity 64.3%; Pred. No. 2.1e-51;

Matches 545; Conservative 0; Mismatches 265; Indels 38; Gaps 11;

```
QY 221 CCCGAGAAACCTGATACAGTGAAGAAAGTGATGTTGTTCTCTCGACATCTACT 280
DB 158572 CTCACAAAAAAGTTGATTCCTCTTCTACACAGACACTT 158513
QY 281 TCACCTCAATATCTATGTTACTTACAGCAATTAATCAGGCCAACCTTAAGTTT 340
DB 158512 AAATTTCTAATTTCTCTATGTAATGTAATGAACTTAATCAGTGAATTTT 158453
QY 341 TATTTGTATAGTATTTTGTGTTAAGCAGACATGTCGTCGAGAAATGATGTTG 400
DB 158452 GAATATCCAAAGTCAGATTTTAAAGA-ATAAGGTCAGAGCTCAG-AACTGATGTTT 158395
QY 401 GTTACTTGATTTAATA-----TATCAACTGGGTTAAATAAAA-----AA 443
DB 158394 CTAACTGACTTAATTAACAGATCTTCTATCCAACTTCGGAAATGAATCTTTT 158335
QY 444 AATGTCATTACTTAATAAAAAAACCAATPACCAACAGACTTACTTCCCTTGGAAAA 503
DB 158334 ACTTGCTGTGCTGTTAAGATCACTAATAATCAATCTATTTTCCCTCGAANA 158275
QY 504 GCACATTTACAAAGG---CTGATCTTACGCTTATATTTACATAAAGAAATTAACCA 560
DB 158274 GCACAGTTTACAGAGCCAACTCTCTCTTATTTGTATACAAATTAACCGA 158215
QY 561 GGTCCCGATATACCTGTAATTTATTCCTA-----AAAGACAGAACTTCACTATG 613
DB 158214 AGTCCGACCTATTTGTATATTTTATTCCTAAGGAAAAACAGAACTTCACTATG 158155
QY 614 CTTTAAATTTAAGTATACCTCAGATACCTGCAAGCTTGAAGCTTACCAAGCAGG-- 671
DB 158154 CTTCAACATTAAGTATTAATTAATCTCAGATATTTTCCAGCTTACACGCAAAATCA 158095
QY 672 ACAGACACAGAGCCCA--GCTACTCTTAGGAAA---TAAATGAGGTGCTCTCTGA 726
DB 158094 TCAGACAAAGAGATCAACTGCTCTCTAGAGAAATCACTTAATGCGGTGCTT 158035
QY 727 CATGCCAGGGGCTTTGCGTGTGCTG--TGTTCAGAAAGCTTTCAGATACACAGTC 785
DB 158034 AATGCCAGGGGCTCTGTAACAGATCGTATTTTCCAGAGGGTTCGTGAGCATGGGTC 157975
QY 786 CTTTGGGCGATTAAGACATATGCTGATTTTCCGTTGCTGCTAGTGAACCCCTAC 845
DB 157974 TGGTTGAGAGGCGAGCATTTCTGATTTTCTGTTGCTGCTGCTAGTGAACCCCTAC 157915
QY 846 AGAAGATAGTGGGTGAGCCAGG--GGGCGAGAGCGGTGCTGACATGTCGTGCTCT 904
DB 157914 AGAAGATTAACGCTAAGCCAGAGAGGCGGAGAGCCACTACATGCTGCTGCTCT 157855
QY 905 TATCACTTATCATATTAAGGAAAGAAAGTATGATTGATTAATCACTGTAACATC 964
DB 157854 TATCACTTATCATATTAAGGAAAGAAAGTATGATTGATTAATCACTGTAACATC 157795
QY 965 AGGAGGAAACAAAGAGCCGTGTCAGACCTCTGAGTGAAGGAGGCTGCTCTCTCTC 1024
DB 157794 TGGGAGAGAGAAACAAAGACCGTGAAGCTGCTCTGTAAGCTGAACAGCAGCCTCC 157735
QY 1025 AGACAGC 1032
DB 157734 AAGTGAGC 157727
```

ID	ABQ81802/c	standard; DNA; 240825 BP.
XX	ABQ81802/c	
AC	ABQ81802;	
XX	14-NOV-2002 (first entry)	
DT	Human PG-3 gene SEQ ID NO:1.	
XX		
DE	Human PG-3; chromosome 8; chromosome 8p23; polymorphic; SNP;	
XX	single nucleotide polymorphism; biallelic marker; DNA repair;	
KW	recombination; cell cycle control; gene; ds.	
KW		
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	5'UTR	1..2000
FT		/*tag= a
FT	primer_bind	1823..1840
FT		/*tag= df
FT	primer_bind	1980..1998
FT		/*tag= dg
FT	misc_binding	1987..2011
FT		/*tag= nh
FT		/note= "probe"
FT	misc_feature	1999
FT		/*tag= ad
FT		/standard_name= "single nucleotide polymorphism (SNP)"
FT		/note= "polymorphic base G or C; the nucleotide is given as an IUPAC ambiguity code in the specification"
FT	primer_bind	complement(2000..2018)
FT		/*tag= dh
FT	exon	2001..2079
FT		/*tag= b
FT		/label= A
FT	intron	2080..4626
FT		/*tag= C
FT		/number= 1
FT	primer_bind	complement(2108..2125)
FT		/*tag= di
FT	primer_bind	4459..4577
FT		/*tag= dj
FT	primer_bind	4582..4600
FT		/*tag= dk
FT	misc_binding	4589..4613
FT		/*tag= nj
FT		/note= "probe"
FT	misc_feature	4601
FT		/*tag= ae
FT		/standard_name= "single nucleotide polymorphism (SNP)"
FT		/note= "polymorphic base A or G; the nucleotide is given as an IUPAC ambiguity code in the specification"
FT	primer_bind	complement(4602..4620)
FT		/*tag= dl
FT	exon	4627..4718
FT		/*tag= d
FT	intron	4719..10114
FT		/*tag= e
FT		/number= 2
FT	primer_bind	complement(4891..4908)
FT		/*tag= dm
FT	primer_bind	10007..10025
FT		/*tag= dn
FT	exon	10115..10233
FT		/*tag= f
FT		/label= C
FT	primer_bind	10209..10227
FT		/*tag= do
FT	misc_binding	10216..10240
FT		/*tag= nj
FT		/note= "probe"

FT	misc_feature	10228	/tag= af	/standard_name= "single nucleotide polymorphism (SNP)"	/note= "polymorphic base G or T; the nucleotide is given as an IUPAC ambiguity code in the specification"	complement(10229..10247)
FT	primer_bind	10234..26809	/tag= dp			
FT	intron	10234..26809	/tag= g			
FT	primer_bind	10267..10285	/number= 3			
FT	misc_binding	10274..10298	/tag= dq			
FT		10274..10298	/tag= nk			
FT	misc_feature	10286	/note= "probe"			
FT		10286	/tag= ag			
FT			/standard_name= "single nucleotide polymorphism (SNP)"			
FT			/note= "polymorphic base G or T; the nucleotide is given as an IUPAC ambiguity code in the specification"			
FT	primer_bind	10287..10305	/tag= dr			
FT	misc_feature	10370	/tag=			
FT		10370	/tag= ah			
FT			/standard_name= "single nucleotide polymorphism (SNP)"			
FT			/note= "insertion of G; the nucleotide is given as an IUPAC ambiguity code in the specification"			
FT	primer_bind	26810..26897	/tag= ds			
FT	exon	26810..26897	/tag= h			
FT			/label= D			
FT	intron	26898..31356	/tag= i			
FT		26898..31356	/number= 4			
FT	exon	31357..31471	/tag= j			
FT		31357..31471	/label= E			
FT	intron	31472..34260	/tag= k			
FT		31472..34260	/number= 5			
FT	exon	34261..34404	/tag= l			
FT		34261..34404	/label= F			
FT	intron	34405..37376	/tag= m			
FT		34405..37376	/number= 6			
FT	exon	37377..37466	/tag= n			
FT		37377..37466	/label= S			
FT	intron	37467..39703	/tag= o			
FT		37467..39703	/number= 7			
FT	primer_bind	39556..39574	/tag= dt			
FT	exon	39704..40858	/tag= p			
FT		39704..40858	/label= T			
FT	primer_bind	39877..39896	/tag= du			
FT	primer_bind	39925..39943	/tag= dv			
FT	misc_binding	39932..39956	/tag= nl			
FT		39932..39956	/note= "probe"			
FT	misc_feature	39944	/tag= ai			
FT		39944	/standard_name= "single nucleotide polymorphism (SNP)"			
FT			/note= "polymorphic base G or T; the nucleotide is given as an IUPAC ambiguity code in the specification"			
FT	primer_bind	39945..39963	/tag= dw			
FT	primer_bind	39953..39970				
FT	primer_bind					

```

FT primer_bind /+tag= dx
FT 39954. .39972
FT /+tag= dy
FT 39961. .39985
FT /+tag= nm
FT /note= "probe"
FT 39973
FT /tag= a_j
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "polymorphic base G or C; the nucleotide is given
FT as an IUPAC ambiguity code in the specification"
FT primer_bind /+tag= dz
FT /complement(40242. .40259)
FT /+tag= ea
FT 40859. .50435
FT /+tag= g
FT /number= 8
FT 41137. .41154
FT /+tag= eb
FT 41366. .41384
FT /+tag= ec
FT 41373. .41397
FT /+tag= nm
FT /note= "probe"
FT 41385. .41403
FT /+tag= ed
FT 41385
FT /tag= ak
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "polymorphic base A or C; the nucleotide is given
FT as an IUPAC ambiguity code in the specification"
FT primer_bind /+tag= ee
FT 41392. .41416
FT /+tag= no
FT /note= "probe"
FT 41404
FT /tag= a_l
FT /standard_name= "single nucleotide polymorphism (SNP)"
FT /note= "polymorphic base A or C; the nucleotide is given
FT as an IUPAC ambiguity code in the specification"
FT primer_bind /+tag= ef
FT /complement(41564. .41581)
FT /+tag= eg
FT 42122. .42141
FT /+tag= eh
FT 42213. .42231
FT /+tag= ei
FT 42220. .42244
FT /+tag= np
FT /note= "probe"
FT 42232
FT /tag= am
FT /standard_name= "single nucleotide polymorphism (SNP)"

Query Match 19.5%; Score 238; DB 6; Length 240825;
Best Local Similarity 64.3%; Pred. No. 2, 1e-51;
Matches 545; Conservative 0; Mismatches 265; Indels 38; Gaps 11;

QY 221 CCCGAGAAACCTGTATACAGTAAAGGATGATGTTCTTCTCGACATACT 280
DB 158572 CTCAAAAAAGGTTGATTCCTCTTCTACACAGACACTT 158513
QY 281 TCACCTAAATATCTTACTTAACAGACATTAATCGGCCAACACTTAAGTTT 340
DB 158512 AAATTCCTAATCTTATGTAATGAATGAACCTTAATACAGTATTTTGAAGT 158453
QY 341 TATTTGTATAGTATTTTGTGTTAAGCACAGACATGTGATGAGAAAAGTATGTTG 400
DB 158452 GAATATCCAGTGAAGTATTTTAAAGA-ATAAAGGTCAAGCTCAG-AAACTGATGTTT 158395

```

```

QY 401 GTAACCTTGATTTAATA-----TATCAACTGGGTTAAATAAAA-----AA 443
DB 158394 CTAACTGACTTAATAACAGATCTTCTATCCAACTGTCGGAATGAATAATCTTTGTA 158335
QY 444 AATGTCATTAACCTTAATAAAAAAAAAAACCAATACCAACAAGACTTACTCCCTCGAAAA 503
DB 158334 ACTTGCTAGTGGTCTCAAGAAATCACTAAATCAAGAAATCTAATTTTCCCTCGAAAAAG 158275
QY 504 GCACATTTACAAAGG---CTGATCTTAGCCTTTATATTAACATAAAGAAATAAACCA 560
DB 158274 GCACAGTTTACAGAGGCCAACTTCTCTCTTTATTTGTATAACAAAAATAAACCGA 158215
QY 561 GGTCCCGATATAGCTGTAAATTTATTCCTA-----AAGAACAGAACTTTCACTATG 613
DB 158214 AGTCCGACCTAATTTGTATATTTTATTCCTAAAGAAAAAACAGAACTTTCAATGTA 158155
QY 614 CTTTAAATTAATTAAGTATACCTCGATATCTGCAAGCTTGGCCCAACAGAGAG-- 671
DB 158154 CTTCAACATTAAGTATTAATTAACCTCAGATATTTTGCAGCTTAGCAGCAAAAATCAGTT 158095
QY 672 ACAGACAAGAGAGCCCA-GCTACTCTAGAGAA-----TAATTAAGGTGTGCTCTGA 726
DB 158094 TCAGACAAAAGAGATTAACCTGCTCTCTCTAGAGAAATTAATTTGGGTGTGCTTAGGA 158035
QY 727 CATGCCAGGGGCTTTGTGGCTGTGCTG-TGTTCCAGAAAGCTTGTGAGTACACAGTC 785
DB 158034 AATGCCAGGGGCTGTGTAACAGATCGTGTTCACAGAGGGTTCGACAGCATGGGTC 157975
QY 786 CTTTGGGCGATTAAGACATATGCTGTGATTTTTCCTGCTGCTAGTACCCTCTAC 845
DB 157974 TGGTTGGAAGGAGAGGATTTCTCTGATTTTTCCTGTTCCCTGCTAGTACCCCTTAC 157915
QY 846 AGGAAGATATAGTGGTATGAGCCAGG-GGGCGAGAGGCTGCTGACATGTCTGCTCTCT 904
DB 157914 AGGAAGATTAACGCTTAAGCCAGAGAGGCGGAGCCGACATGATGTCTGCTCTCT 157855
QY 905 TATCAACTTATCATATTAAGGAGAAAGAAAGTATGATTCGATCTGACACTGTAGACTC 964
DB 157854 TATCAACTTATCATATTAAGGAGAAAGAAAGTATGATTCGATCTGACACTGTAGACTC 157795
QY 965 AGGGGAGAAACAAAGTCCGTCAGACACTCTGTGATGAGAGGGTGTCTTCTCTTC 1024
DB 157794 TGGGAGAGAGAGAAACAAAGACCTGTGAAGCTGCTCTGTAAAAAGCTGACACAGCCCTCC 157735
QY 1025 AGGACAGC 1032
DB 157734 AAGTGAGC 157727

RESULT 6
ADQ09378
ID ADQ09378 standard; DNA; 2424 BP.
XX
AC ADQ09378;
XX
XX
XX 09-SEP-2004 (first entry)
XX
XX Murine Angiopoietin-2 DNA.
DE
XX
XX Mouse; Angiopoietin-2; gene; de; antisense oligonucleotide;
XX phosphotriphosphate linkage; 2'-O-methoxyethyl sugar moiety;
XX 5-methylcytosine; hyperproliferative disorder; cancer; cytosolic.
XX
XX Mus musculus.
XX
XX
XX Key Location/Qualifiers
XX CDS 211..1701
XX FT /+tag= a
XX FT /product= "Murine Angiopoietin-2"
XX
XX US2004115640-A1.
XX
XX 17-JUN-2004.
PD

```





CC i.e. a phosphorothioate linkage, at least one modified sugar moiety,  
CC preferably a 2'-O-methoxyethyl sugar moiety, or at least one modified  
CC nucleobase comprising a 5-methylcytosine. The antisense compounds are  
CC useful for modulating the expression of the human Angiopoietin-2  
CC polypeptide and in preparation of a composition for treating  
CC hyperproliferative disorders, e.g. cancer. This sequence represents DNA  
CC encoding a human Angiopoietin-2 polypeptide of the invention.

SO Sequence 62705 BP; 19403 A; 12918 C; 12719 G; 17665 T; 0 U; 0 Other;

Query Match 17.4%; Score 211.8; DB 12; Length 62705;  
Best Local Similarity 74.3%; Pred. No. 1e-44;  
Matches 335; Conservative 0; Mismatches 107; Indels 9; Gaps 5;

QY 591 AAAGAACGAACTTTTACTATGCTTTAAATTAAAGTATTACTCAGATCTGCA 650  
DB 31 AAAAAACGAACTTTTACTATGCTTTAAAGTATTACTCAGATCTGCA 90  
QY 651 GCTTAGCGCTCAAAACGAGCAG--ACAGACAAACAGAGCCCA-GCTACTCTCTAGGAA-- 705  
DB 91 GCTTAGCAGCGCAAAATTCAGTTTCAGACAAAGATCAACTGCTCTCTAGGAAATA 150  
QY 706 --TAATTAGGAGTGCTGCTCTGACATGCCCCAGGGGCTTGTGGCTGCTCG--TGTCCCA 762  
DB 151 CTAAATTGGGGTGCTGCTCTAGGAAATGCCAGGGGCTCTGTACAGATCGCTTTTCCCA 210  
QY 763 GAAGGCTTCTGCGATACACAGTCTTTGGGGCAGTAAGCACTATGCTCTGATTTTCCCG 822  
DB 211 GAGGGTTTCTGCGACATGGGCTCTGTTGAGGGCAGGCACTTCTGATTTTCCCG 270  
QY 823 TTGCTGCTAGTACAGCCCTACAGAGATAGTGGCTAGGCAAG--GGCGGACCGGT 881  
DB 271 TTGCTGCTAGTACAGCCCTACAGAGATAGTGGCTAGGCAAG--GGCGGACCGGT 330  
QY 882 GGCTGACATGCTGCTGCTCTTATCACTTATCATATATAGGAAAGTATGAT 941  
DB 331 CACTACACATGCTGCTGCTCTTATCACTTATCATATATAGGAAAGTATGAT 390  
QY 942 TCGATATCTGACATCTAGTACCTCAGGGGAGAAACAAAGTCCGTCAGACCTCTGAGT 1001  
DB 391 TCGATATCTGACATCTAGTACCTCAGGGGAGAGAAAGAACCGTGAAGCTGCTCT 450  
QY 1002 GAGCAGGGGCTGCTCTCTCTCTCAGACAGC 1032  
DB 451 GTAAAAAGTGACACAGCCCTCCCAAGTGAGC 481

RESULT 10  
ABX63024  
ID ABX63024 standard; cDNA; 3251 BP.  
XX  
AC ABX63024;  
XX  
DT 25-FEB-2003 (first entry)  
XX  
DE Human cDNA #24 differentially expressed in activated vascular tissue.  
XX  
KW Human; gene; ss; vascular tissue; cytosolic; atherosclerosis; cardiant;  
KW hypotensive; antidiabetic; gynaecological; vasotropic; cerebroprotective;  
KW gene therapy; vascular disease; cancer; coronary; artery disease;  
KW hyperension; diabetes; pre-eclampsia; restenosis;  
KW ischaemia-reperfusion injury; stroke.  
XX  
OS Homo sapiens.  
XX  
PN US2002137081-A1.  
XX  
PD 26-SEP-2002.  
XX  
PE 08-JAN-2002; 2002US-00044090.  
XX  
PR 28-JUL-2000; 2000US-0222469P.  
PR 08-JAN-2001; 2001US-0260483P.

XX  
PA (BAND/) BANDMAN O.  
XX  
PI Bandman O;  
XX  
DR WPI; 2003-110597/10.  
XX  
PT Combination for diagnosing, staging, treating, or monitoring the  
PT progression of treatment of a vascular disease, e.g. atherosclerosis,  
PT comprises several cDNAs that are differentially expressed in activated  
PT vascular tissue.  
XX  
PS Claim 1; Page; 18pp; English.

CC This invention relates to a combination comprising several cDNAs that are  
CC differentially expressed in activated vascular tissue. The invention also  
CC discloses a high throughput method for detecting differentially expressed  
CC cDNAs in a sample. The cDNAs of the invention may have  
CC antiarteriosclerotic; cytosolic; cardiant; hypotensive; antidiabetic;  
CC gynaecological; vasotropic and cerebroprotective activities and may be  
CC used in gene therapy. The cDNAs of the invention may be used in a high-  
CC throughput methods for detecting differential expression of one or more  
CC cDNAs in a sample, or screening several molecules or compounds to  
CC identify a molecule or compound that specifically binds a cDNA of the  
CC invention. A protein encoded by the cDNA may be used to screen several  
CC molecules or compounds to identify a ligand that specifically binds to  
CC the protein, or to produce or purify an antibody to the protein that can  
CC be used to detect a protein in a sample or purify a natural or  
CC recombinant protein from a sample. The nucleotides may be useful for  
CC diagnosing, staging, treating, or monitoring the progression of treatment  
CC of a vascular disease, e.g. atherosclerosis, cancer, coronary artery  
CC disease, hypertension, diabetes, pre-eclampsia, ischaemia-reperfusion  
CC injury, restenosis, or stroke. The cDNAs can also be used for large-scale  
CC genetic or gene expression analysis of several new nucleic acid  
CC molecules. Antibodies to the proteins encoded by the cDNAs are useful for  
CC diagnosing pre-pathologic disorders, and chronic or acute diseases  
CC associated with abnormalities in the expression, amount or distribution  
CC of the protein. The present sequence represents a cDNA of the invention  
CC that is differentially expressed in activated vascular tissue. Note: The  
CC sequence data for this patent did not form part of the specification, but  
CC was obtained in electronic format directly from USPTO at  
CC <http://seqdata.uspto.gov/sequence.html?DocID=20020137081>  
CC  
XX

SO Sequence 3251 BP; 1083 A; 645 C; 674 G; 849 T; 0 U; 0 Other;

Query Match 12.8%; Score 156.6; DB 8; Length 3251;  
Best Local Similarity 74.9%; Pred. No. 1.2e-30;  
Matches 209; Conservative 0; Mismatches 69; Indels 1; Gaps 1;

QY 755 TGTTCACAGAGGCTTCTGACATACAGTCTTTGGGACAGTAAGCACTATGCTCTGAT 814  
DB 12 TTTTCCAGAGGGGTTTCTGACAGATGGTCTGTGTGAGGGCAGCACTTGTCTGTAT 71  
QY 815 TTTTCCGTTGCTGCTAGTACCCCTACAGAAAGATAGTGGTGAAGCCAG--GGCG 873  
DB 72 TTTTCTGTGCTGCTGCTAGTACCCCTACAGAAAGATAGTGGTGAAGCCAGGGGCG 131  
QY 874 GAGCGGCTGCTGACATGCTGCTCTTATCACTTATCATATATAGGAAAGAA 933  
DB 132 GAGCAGCCACACTACACATGCTGGGCTCTTATCACTTATCATATATAGGAAAGAA 191  
QY 934 TGATTGATTCGATACGATCTGAGACTCAGGGGAGAAACAAAGAGTCCGTGAGACC 993  
DB 192 TGATTGATTCGATACGATCTGAGACTCAGGGGAGAGAAACAAAGAGCCGTGAAA 251  
QY 994 TCTGAGTGAGCAGGCTGCTCTCTCTCTCAGACAGC 1032  
DB 252 GCTGCTCTGTAAAGCTGACACAGCCCTCCCAAGTAGAC 290

RESULT 11  
ADD69421  
ID ADD69421 standard; DNA; 763 BP.



Db 1770 AGATCGGTTTTTTTATAGAGGTTTTTGTATGAGGTTTTGTGAGGGTAGGTAATTTT 1829  
QY 807 GCTCTGATTTTTTCTGTTGCTGGCTAGTGACCCCCACAGAAAGATAGGGGTAGCCCA 866  
Db 1830 GTTTTGATTTTTTTTGTGTTGTTGGTTAGTGATTTTTTTATAGAAAGATACGGTTAAGTTA 1889  
QY 867 GG-GGGGGAGCGGCTGGCTGCACATGCTGGCTGCTTATCAACTATCATATAAGG 925  
Db 1890 GGAAGGGCGAGTAGTTATATATATATGTTGGTTGTTTTTATTAATTATATATAAGGA 1949  
QY 926 AAGGAAGTATGATTCGATCTGACACTGTAGATCTCAGGGAGAAACAAGA 980  
Db 1950 AAGGAAGTATGATTCGATCTGACACTGTAGATCTCAGGGAGAGAGAAATA 2004  
RESULT 13  
ID ABQ67158/c  
ID ABQ67158 standard; DNA: 2846 BP.  
AC ABQ67158:  
XX  
DT 28-AUG-2002 (first entry)  
XX  
DE Human angiogenesis associated polynucleotide SEQ ID NO 188.  
XX  
KM Human: angiogenesis; methylation; eye disease; glaucoma; tumour;  
KM inflammation; rheumatoid arthritis; diabetic retinopathy; antileucers;  
KM macular degeneration; inflammatory bowel disease; Crohn's disease;  
KM antiathermatic; antiatheritic; antidiabetic; antipsoriatic;  
KM antiarteriosclerotic; ds.  
XX  
OS Homo sapiens.  
XX  
PN WO20024654-A2.  
PD 13-JUN-2002.  
XX  
PF 06-DEC-2001; 2001WO-EP014320.  
XX  
PR 06-DEC-2000; 2000DE-01061338.  
PA (EPIC-) EPIGENOMICS AG.  
PI Schacht O;  
XX  
DR WPI: 2002-500450/53.  
PT New nucleic acid fragments from chemically treated angiogenesis-  
PT associated genes, useful for determining methylation status, e.g. in  
PT diagnosis or treatment of cancer.  
XX  
PS Claim 1; SEQ ID NO 188; 41pp + Sequence Listing; German.  
XX  
CC The invention relates to a nucleic acid (I) comprising a segment of 18  
CC bases of chemically pre-treated DNA of angiogenesis-associated genes (II)  
CC having sequences (ABQ66971-ABQ67178) or their complements. (I), also  
CC related oligomers, are used to evaluate the methylation status and/or  
CC single-nucleotide polymorphisms, in angiogenesis-related genes, for  
CC diagnosis and treatment of eye diseases, proliferative retinopathy,  
CC neovascular glaucoma, solid tumours, inflammation, rheumatoid arthritis,  
CC diabetic retinopathy, macular degeneration caused by neovascularisation,  
CC psoriasis, arteriosclerosis, inflammatory bowel diseases, ulcers and  
CC Crohn's disease. Note: The sequence data for this patent did not form  
CC part of the printed specification, but was obtained in electronic format  
CC directly from WIPO at ftp.wipo.int/pub/published\_pat\_sequences  
XX  
SQ Sequence 2846 BP; 759 A; 45 C; 565 G; 1477 T; 0 U; 0 Other;

Db 1651 CGACAAACGAAACCTCCGCTCCAAAAAATTTATATATTCCTCTT 1592  
QY 266 TCCCTCGACATATCTCACTCAATATCTAATGTTACTTTAACAAGCAATATACAGGCA 325  
Db 1591 TCTACCAAAACCTTAATTTCTAATTTCTATATATATATATTAACCTTAATCAATACA 1532  
QY 326 AACCACTTAAGTTTATATTTGATATGATATTTTGTATAGGACAGACATGTGAGCTG 385  
Db 1531 ATATATTTTAAAAATTAATTAATCAATTAATTAATTTTAAAAAATTAATTAATCAAT 1474  
QY 386 AGAAAACTGATGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTT 435  
Db 1473 CAAAAACCTAATTTTCTAAGCTAATTAATTAATTAATTAATTTTCTAATTAATTCGAAA 1414  
QY 436 TAAAA-----AAAAATGTCATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 488  
Db 1413 TAAAAATCTTTTATATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1354  
QY 489 CTTCCTTGGAAAAAGCA---TTTACAGGCTGATCTTACCTTTATTTTACATA 545  
Db 1353 TTTCCCTTAATAAAACACAGCTTTACAAAAACCAACTTCTCTTTATTTATATA 1294  
QY 546 AAGAAATTAACCAAGTCCGATATAGCTGTAATTTATTTCT------AAAGACA 598  
Db 1293 ACAAATTAACCGAAATCTTAACCTTATATATTTTATTTTATTTTATTTTATTTTATTT 1234  
QY 599 GAACTTTCACTATGCTTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 657  
Db 1233 AAACTTCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1174  
QY 658 CTACAAACGAGCAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGAC 711  
Db 1173 CGACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1114  
QY 712 GGGTGTGCTCTGACATGCCCAGGGGTCTGTGCTGTGTG-.-TTCCAGAGGCTT 770  
Db 1113 AATATATACCTTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1054  
QY 771 CTGCAATACAGTCTTTGGGGCAGTAAGCACTATGCTGATTTTCTGTGCTGTG 830  
Db 1053 CTACACATTAATTTCTTAATTAATTAATTAATTAATTAATTAATTAATTTTCTATTA 994  
QY 831 CTAGTGAACCCCTACAGAGATAGTGGTGAGCC-AGGGGGGAGAGGCTGGCTGCAC 889  
Db 993 CTAAATACCCCTACAAAAAATTAAGACTTAACCAAAAAAAGGAAACCACTACAC 934  
QY 890 ATGCTGCTGCTCTTATCACTTATCATTTAAGGAGAAAGTATGATTCGATAC 949  
Db 933 ATATCTAATCTATCTTATCACTTATCATTTAATTAATTAATTAATTTGAAATAC 874  
QY 950 TGACACTGTAGCTCAGGGAGAAACAAGA 980  
Db 873 TAACATATTAATCTTAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTA 843  
RESULT 14  
ID ADA71938/c  
ID ADA71938 standard; DNA: 2000 BP.  
XX  
AC ADA71938:  
XX  
DT 20-NOV-2003 (first entry)  
XX  
DE Rice gene, SEQ ID 5263.  
XX  
KM Plant; bacterial infection; fungal infection; viral infection; rice;  
KM gene; ds.  
XX  
OS Oryza sativa.  
XX  
PN WO200300898-A1.  
XX



PD 03-JAN-2003.  
 XX  
 PF 22-JUN-2001; 2001WO-IB001105.  
 XX  
 PR 22-JUN-2001; 2001WO-IB001105.  
 XX  
 PA (SYGN) SYNGENTA PARTICIPATIONS AG.  
 XX  
 PI Chang H, Chen W, Cooper B, Glazebrook J, Goff SA, Hou Y,  
 PI Katagiri F, Qian S, Tao Y, Whitlam S, Xie Z, Zhu T, Zou G;  
 DR WPI; 2003-175290/17.  
 XX  
 PT Identifying at least one gene involved in plant resistance or response to  
 PT pathogenic infection for conferring resistance or tolerance to a plant to  
 PT bacterial, fungal or viral infection by determining or detecting plant  
 PT gene expression.  
 XX  
 PS Claim 27; SEQ ID NO 5263; 899pp; English.  
 XX  
 CC The present invention relates to a method (M1) for identifying genes  
 CC involved in plant resistance or response to pathogenic infection. M1  
 CC comprises identifying a gene whose expression is significantly altered in  
 CC the incompatible interaction of plant gene expression relative to  
 CC expression of the gene in an uninfected plant, in a mutant plant that  
 CC does not express a gene associated with response to pathogenic infection,  
 CC or in a corresponding incompatible or compatible interaction. (M1) is  
 CC useful for conferring resistance to resistance or tolerance to a plant to  
 CC bacterial, fungal or viral infection. The present sequence was used to  
 CC illustrate the invention.  
 XX  
 SQ Sequence 2000 BP; 336 A; 265 C; 284 G; 363 T; 0 U; 752 Other;  
 XX  
 Query Match 4.8%; Score 58; DB 8; Length 2000;  
 Best Local Similarity 11.6%; Pred. No. 0.00013;  
 Matches 49; Conservative 204; Mismatches 169; Indels 2; Gaps 1;  
 XX  
 QY 329 CACCTTAAGTTTATTTAGTATGATTTGTTGTTAAGCAGACATGAGTCTGAGA 388  
 DB 1009 SMSWYVAMWKYTKWTATTAISSTWKWYMAKRAYAMRSRKTWCTGGRMATYCGTKRW 950  
 QY 389 AAGAGATGTGTGTAATCTGATTTAATAATCAAACTGGGTTAAATAAAAAAATGT 448  
 DB 949 AAGRRMRMAACWCCMKMKWKTSCMKWKYRTSCWTTMMGAKRYAYVAMRRRTY 890  
 QY 449 GCATACTTAAAAAACAACCAATACCAAGACTTACTTCCCTTGAAGAACACA 508  
 DB 889 KMSWRMTWTMTKMAVTMTTCMAKMYVAITGMATMMWRMYTTCYVMTCAKCKYKMA 830  
 QY 509 TTTACAAGGCTGATCTTAGCCTTATATTTACAATAAAGAAAATAAACCAAGTCCGA 568  
 DB 829 MTKWMTTACAWRAISWRRAAGMKRKYMKKRAVAMWMBRCMKAGMARSRMRK 770  
 QY 569 TATAGCTGAATTTATTTCTTAAGAACAGAACTTTACATATGTTTAAATTAAGT 628  
 DB 769 KXIAIRYVKKMAATMMWMSWRKSYRMWSGMRMWSAMRYCSRKCAKTYVASSART 710  
 QY 629 GATTACCTGATCTGCAAGCTTAGCTTCAAC--GAGCAGACAGACAGAGCC 686  
 DB 709 KRAKSYRRRRMYWKRMWTRRYRWRSCMTBRAMSKRRKAGAGSMSCMMYRGAAS 650  
 QY 687 CCAAGTACTCTTAGAATAATATAGAGGTGGCTCTGACATGCCCAAGGGGTCTTGGG 746  
 DB 649 MWSYKSSAKCKKTRVMTSYMTGYMWSYSSKMSMTSKMSYMGKMTCTMYTSMG 590  
 QY 747 CTGG 750  
 DB 589 STR 586

RESULT 15  
 ABL32893/c  
 ID ABL32893 standard; DNA; 10710 BP.

XX  
 AC ABL32893;  
 XX  
 DT 26-MAR-2002 (first entry)  
 XX  
 DE Human immune system associated gene SEQ ID NO: 866.  
 XX  
 KW Human; immune system disease; cytosine methylation; antiasthmatic;  
 KW antiarteriosclerotic; antianaemic; cyostatic; noctropic;  
 KW neuroprotective; anti-HIV; anticonvulsant; ophthalmological;  
 KW antineumatic; antiarthritic; antidiabetic; antipsoriatic;  
 KW antiinflammatory; cancer; eye disease; arteriosclerosis; anaemia;  
 KW acute myeloid leukaemia; Alzheimer's disease; AIDS; epilepsy;  
 KW neurofibromatosis; rheumatoid arthritis; psoriasis; bowel disease; gene;  
 KW ds.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200200928-A2.  
 XX  
 PD 03-JAN-2002.  
 XX  
 PF 02-JUL-2001; 2001WO-EP007537.  
 XX  
 PR 30-JUN-2000; 2000DE-01032529.  
 XX  
 PR 01-SEP-2000; 2000DE-01043826.  
 XX  
 PA (EPIC-) EPIGENOMICS AG.  
 XX  
 PI Olek A, Piepenbrock C, Berlin K;  
 XX  
 DR WPI; 2002-130909/17.  
 XX  
 PT Nucleic acid comprising fragment of chemically modified gene, useful for  
 PT diagnosis and treatment of diseases associated with abnormal cytosine  
 PT methylation.  
 XX  
 PS Claim 1; SEQ ID NO 866; 32pp + Sequence Listing; German.  
 XX  
 CC The present invention provides a number of human immune system associated  
 CC genes which are modified by the methylation of cytosines. The sequences  
 CC can be used in the diagnosis and treatment of immune system disorders,  
 CC including eye diseases such as retinopathy, neovascular glaucoma and  
 CC macular degeneration, arteriosclerosis, anaemia, cancer, acute myeloid  
 CC leukaemia, Alzheimer's disease, AIDS, epilepsy, neurofibromatosis,  
 CC rheumatoid arthritis, psoriasis and inflammatory/ulcerative bowel  
 CC diseases. The present sequence is a gene of the invention  
 XX  
 SQ Sequence 10710 BP; 3653 A; 81 C; 1963 G; 5013 T; 0 U; 0 Other;  
 XX  
 Query Match 4.5%; Score 55; DB 6; Length 10710;  
 Best Local Similarity 48.8%; Pred. No. 0.0018;  
 Matches 148; Conservative 0; Mismatches 155; Indels 0; Gaps 0;  
 XX  
 QY 256 TATGTTGTTCTTCGACATATCTTCACTCAATATCTATGTTACTTAACAGACATT 315  
 DB 5616 TATAATTTCTAATCTCAGCAATATCCTTATTAATAAATAATTAATCAAAACAACT 5557  
 QY 316 AATCAGGCAACCACTTAAGTTTATTTGTAATGTAATTTGTTAAGCAGACAT 375  
 DB 5556 AATATTTTAATTTCTTTTCAATACATTTATGATTTTATATATATCTTACATATT 5497  
 QY 376 GTGAGTGTGAGAAACTGATGTTGTAACCTGATTTAATATATCAAACTGGTTAAA 435  
 DB 5496 TTAATTAATAATAATAATAAATAATTTAATAATATATATATTAATCTATTTATA 5437  
 QY 436 TAAAAAATAATGTGATTAATCTTAATAAATAAATAAATAAATAAATAAATAAATA 495  
 DB 5436 AATATCTTTCTAACAATACATTAACAAAAATAAATAAATAAATAAATAAATAAATA 5377  
 QY 496 TTGAAAAAGACATTTACAGGGCTGATCTTATGCTTATATATTTCAATTAAGAAATTA 555  
 DB 5376 TAAATTAACAAAATAAATAACCAATAAATAAATAAATAAATAAATAAATAAATAAATA 5317

QY	556	ACC	558
Db	5316	AAC	5314

Search completed: December 12, 2005, 12:16:44  
Job time : 762 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 12, 2005, 11:45:12 ; Search time 4725 Seconds

(without alignments)  
12080.474 Million cell updates/sec

Title: US-10-627-075-1

Perfect score: 1220  
1 gagctctccagaccgacac.....ccaaaaagagagagagatcg 1220

Scoring table: IDENTITY NUC  
Gapop 10.0, Gapext 1.0

Searched: 41078325 segs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

EST:\*  
1: gb\_est1:\*  
2: gb\_est2:\*  
3: gb\_est3:\*  
4: gb\_hic:\*  
5: gb\_est4:\*  
6: gb\_est5:\*  
7: gb\_est6:\*  
8: gb\_est7:\*  
9: gb\_gsa1:\*  
10: gb\_gsa2:\*  
11: gb\_gsa3:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysts of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	276	22.6	699	5	BY716819 BY716819
2	276	22.6	2475	4	AK019860 Mus muscu
3	274.4	22.5	792	7	CNS539462 UT-M-HUO-
4	263	21.6	330	1	BB847613 BB847613
5	253	20.7	354	1	AA389106 mp24h01.r
6	242	19.8	243	4	AK048622 Mus muscu
7	240.4	19.7	299	2	BB649745 BB649745
8	217.8	17.9	912	2	BE306402 601102294
9	208.6	17.1	276	2	BB569161 BB569161
10	204	16.7	665	2	BB612537 BB612537
11	203.4	16.7	661	9	AZ984331 AZ984331
12	185.8	15.2	793	11	CR905089 Sus scrofa
13	178.6	14.6	726	5	BX506958 DKFZp779M
14	175.8	14.4	204	2	BB601394 BB601394
15	167.8	13.8	525	5	BUS60281 BK77609.Y
16	149	12.2	921	6	CB203297 CB203297
17	143.4	11.8	482	8	DN994582 DN994582
18	133.6	11.0	477	2	BE292322 601084072
19	131.2	10.9	592	3	BP320006 BP320006
20	121.8	10.0	466	6	CB548076 AMGNMNC:C
21	119.6	9.8	679	2	BB614741 BB614741
22	93.8	7.7	247	2	BF085046 PMO-GN001

23	58.4	4.8	804	7	CO739081
24	58.4	4.8	819	7	CO739225
25	58.4	4.8	834	7	CO735207
26	57.6	4.7	532	1	AI882255
27	55.8	4.6	454	1	AV682744
28	55.4	4.5	1101	10	AV682744
29	55	4.5	512	5	BX504641
30	55	4.5	928	10	CNS00DKY
31	54.6	4.5	1101	10	CNS00EVL
32	54.4	4.5	1150	10	AG365432
33	54	4.4	1101	10	CNS0021J
34	53.6	4.4	510	5	BK478886
35	53.2	4.4	1000	10	CNS00CCQ
36	53	4.3	639	11	CNS038CX
37	52.8	4.3	978	10	AG129148
38	52.8	4.3	1043	10	CNS0145P
39	52.6	4.3	1131	11	CNS034FO
40	52.6	4.3	759	11	CNS060XV
41	52.6	4.3	1201	10	CNS0167M
42	52.2	4.3	1085	10	CNS02PW4
43	52.2	4.3	1201	10	CNS0167M
44	52	4.3	331	1	AU037151
45	52	4.3	1101	10	CNS0039V

#### ALIGNMENTS

RESULT 1	BY716819	699 bp	mRNA	linear	EST 17-DEC-2002
LOCUS	BY716819	R16EN full-length enriched, 11 days pregnant adult female ovary and uterus Mus musculus cDNA clone 5031400E18.5, mRNA sequence.			
DEFINITION	BY716819				
ACCESSION	BY716819				
VERSION	BY716819.1	GI:27129936			
KEYWORDS	EST.				
SOURCE	Mus musculus (house mouse)				
ORGANISM	Mus musculus				
REFERENCE	1 (bases 1 to 699)				
AUTHORS	Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nikaido, I., Osato, N., Saito, R., Suzuki, H., Yamakawa, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schonbach, C., Gotohori, T., Baldarelli, R., Hill, D.P., Bult, C., Hume, D.A., Quackenbush, J., Schriml, L.M., Kanapin, A., Matsuda, H., Batalov, S., Beisel, K.W., Blake, J.A., Brad, D., Bruscia, V., Chothia, C., Corbani, L.E., Cousins, S., Dalla, E., Drengani, T.A., Fletcher, C.F., Forrest, A., Frazer, K.S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, J., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I.J., Jarvis, E.D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedzierecki, R.M., King, B.L., Konagaya, A., Kurochkin, I.V., Lee, Y., Lemhard, B., Lyons, P.A., Maglott, D.R., Maltais, L., Marchionni, L., McKenzie, L., Miki, H., Nagashima, T., Naita, K., Okido, T., Pavan, W.J., Perte, G., Pesole, G., Petrovsky, N., Pillai, R., Pontius, J.U., Qi, D., Ramachandran, S., Ravasi, T., Reed, J.C., Reed, D.J., Reid, J., Ring, B.Z., Ringwald, M., Sautin, A., Schneider, C., Semple, C.A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M.S., Teasdale, R.D., Tomita, M., Verardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L.G., Wynnshaw-Boris, A., Yanagisawa, M., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kitahara, T., Kono, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Kawai, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shinagawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E.S., Rogers, J., Birney, E. and Hayashizaki, Y.				
TITLE	Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs				
JOURNAL	Nature 420, 563-573 (2002)				

PUBMED  
COMMENT

1246851  
Contact: Yoshitake Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@gsc.riken.jp, URL: http://genome.gsc.riken.jp/  
Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P.,  
Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F.,  
Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawai, J., Kojima, Y.,  
Komodo, S., Kono, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M.,  
Nomura, K., Numazaki, R., Ohno, M., Ohnato, N., Saito, R., Sakazume, N.,  
Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tagami, M.,  
Takeda, Y., Waki, K., Watabiki, A., Muramatsu, M. and Hayashizaki, Y.  
Direct Submision  
Computational Analysis of Full-Length Mouse cDNAs Compared with  
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new  
genes. Genome Res. 10 (10), 1617-1630 (2000)  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer. Genome Res.  
10 (11), 1757-1771 (2000)  
Computer-based methods for the mouse full-length cDNA  
encyclopedia: real-time sequence clustering for construction of a  
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)  
cDNA library was prepared and sequenced in Mouse Genome  
Encyclopedia Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in RIKEN.  
Division of Experimental Animal Research in Riken contributed to  
prepare mouse tissues.  
Please visit our web site (http://genome.gsc.riken.go.jp) for  
further details.

FEATURES  
source

Location/Qualifiers  
1..699  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="5031400E18"  
/sex="female"  
/tissue\_type="ovary and uterus"  
/dev\_stage="11 days pregnant, adult"  
/lab\_host="DH10B"  
/clone\_1ib="RIKEN full-length enriched, 11 days pregnant  
adult female ovary and uterus"  
/note="Site 1: Salt, Site 2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer [5'  
GAGAGAGAGAGATCCAGAGCTCTTTTCTTTTCTTTTCTTTT 3'], cDNA was  
prepared by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. cDNA went through one round of normalization  
to Rot = 10.0 and subtraction to Rot = 100.0. Second  
strand cDNA was prepared with the primer adapter of  
sequence [5' GAGAGAGAGATCTTGATTTATTTAAATATCCCCCCCCCC  
3']. cDNA was cloned into the XhoI and BamHI sites.  
Vector: a modified pluescript KS(+) after bulk excision  
from Lambda PUC I. Cloning sites, 5' end: SalI; 3' end:  
BamHI."

## ORIGIN

Query Match 22.6%; Score 276; DB 5; Length 699;  
Best Local Similarity 100.0%; Pred. NO. 1.7e-58;  
Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
945 GATCTGACACTGTAGACTCAGGGGAGAAACAAAGAGTCGTCGACACTCTGAGAGAG 1004

Db  
1 GATCTGACACTGTAGACTCAGGGGAGAAACAAAGAGTCGTCGACACTCTGAGAGAG 60  
QY  
1005 CAGGGCTGCTCTTCTCTTCAGAGACGCTCCAGTGTCTCCGGGAGAGAGAGAGA 1064  
61 CAGGCTGCTCTTCTCTCTTCAGAGACGCTCCAGTGTCTCCGGGAGAGAGAGAGA 120  
Db  
1065 GACAGGCACTGGGAAAGAGCTCTGCTCGGAGCGAGAGAGCTTCACTGATGACTTAT 1124  
121 GACAGGCACTGGGAAAGAGCTCTGCTCGGAGCGAGAGAGCTTCACTGATGACTTAT 180  
QY  
1125 CACACGGCAGACGCTGCTGCTTGAAGACAGCTGAGAGCTCAGAGCGCAAGTTGCTGA 1184  
181 CACACGGCAGACGCTGCTGCTTGAAGACAGCTGAGAGCTCAGAGCGCAAGTTGCTGA 240  
QY  
1185 ACTCAGCTTTAGAACCCAAAAGAGAGAGAGATG 1220  
241 ACTCAGCTTTAGAACCCAAAAGAGAGAGAGATG 276

RESULT 2  
AK019860  
LOCUS  
DEFINITION  
AK019860 2475 bp mRNA linear HTC 03-Apr-2004  
Mus musculus 11 days pregnant adult female ovary and uterus cDNA,  
RIKEN full-length enriched library, clone:5031400E18  
product:ANGIOPOIETIN-2 PRECURSOR, full insert sequence.  
ACCESSION  
AK019860 1 GI:12860231  
VERSION  
AK019860  
KEYWORDS  
HTC; CAP trapper.  
SOURCE  
Mus musculus (house mouse)  
ORGANISM  
Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.

REFERENCE  
AUTHORS  
Carninci, P. and Hayashizaki, Y.  
TITLE  
High-efficiency full-length cDNA cloning  
JOURNAL  
Meth. Enzymol. 303, 19-44 (1999)  
PUBMED  
10349636  
AUTHORS  
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K.,  
Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
TITLE  
Normalization and subtraction of cap-trapper-selected cDNAs to  
prepare full-length cDNA libraries for rapid discovery of new genes  
JOURNAL  
Genome Res. 10 (10), 1617-1630 (2000)  
PUBMED  
11042159  
AUTHORS  
Shibata, K., Itoh, M., Aizawa, K., Nagaoaka, S., Sasaki, N., Carninci, P.,  
Kono, H., Akiyama, J., Nishii, K., Katsunai, T., Tashiro, H., Itoh, M.,  
Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A.,  
Yamamoto, R., Watanabe, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K.,  
Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watabiki, M.,  
Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsunaga, S., Kawai, J.,  
Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and Hayashizaki, Y.  
TITLE  
RIKEN integrated sequence analysis (RISA) system-384-format  
sequencing pipeline with 384 multicapillary sequencer  
JOURNAL  
Genome Res. 10 (11), 1757-1771 (2000)  
PUBMED  
11078661

TITLE  
THE RIKEN Genome Exploration Research Group Phase II Team and the  
FANTOM Consortium.  
REFERENCE  
Functional annotation of a full-length mouse cDNA collection  
JOURNAL  
Nature 409, 685-690 (2001)  
PUBMED  
11078661

TITLE  
The FANTOM Consortium and the RIKEN Genome Exploration Research  
Group Phase I & II Team.  
REFERENCE  
Analysis of the mouse transcriptome based on functional annotation  
of 60,770 full-length cDNAs  
JOURNAL  
Nature 420, 563-573 (2002)  
PUBMED  
12475  
AUTHORS  
Adachi, J., Aizawa, K., Akahira, S., Akimura, T., Arai, A., Aono, H.,  
Arakawa, T., Bono, H., Carninci, P., Fukuda, S., Fukunishi, Y.,  
Furuno, M., Hanagaki, T., Hara, A., Hayatsu, N., Hiramoto, K.,











Db 3 GAGTCCGTGACGCTCTGTGAGTCAGACAGGCTGTCTCTCTCAAGACACCTCCGAG 62  
Qy 1039 TGTGCCGGGAGAGAGAGAGAGAGACAGCAGCAGCTGGGAAAGAGCTGTGGGGACGG 1098  
Db 63 TGTGCCGGGAGAGAGAGAGAGAGACAGCAGCAGCTGGGAAAGAGCTGTGGGGACGG 122  
Qy 1099 AGAAGGCTCTCACTGATGCACTTATTTCACAGCGACAGCCCTGTGCTTATGACAGAGCT 1158  
Db 123 AGAAGGCTCTCACTGATGCACTTATTTCACAGCGACAGCCCTGTGCTTATGACAGAGCT 182  
Qy 1159 GAGAGCTCAGAGCAGAGTTTGTCTGAATCTACAGCTTAGAACCCAAAAGAGAGAGAA 1218  
Db 183 GAGAGCTCAGAGCAGAGTTTGTCTGAATCTACAGCTTAGAACCCAAAAGAGAGAGAA 242  
Qy 1219 TG 1220  
Db 243 TG 244

RESULT 8  
BE306402 912 bp mRNA linear EST 26-OCT-2000  
LOCUS 601102294F1 NCI\_CGAP\_Lu29 Mus musculus cDNA clone IMAGE:3494566 5',  
DEFINITION mRNA sequence.  
ACCESSION BE306402  
VERSION BE306402.1 GI:9159558  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Murioidea; Muridae; Murinae; Mus.  
REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>  
1 (bases 1 to 912)  
AUTHORS National Institutes of Health, Mammalian Gene Collection (MGC)  
TITLE Unpublished (1999)  
JOURNAL Contact: Robert Strausberg, Ph.D.  
COMMENT Email: cgabbs-r@mail.nih.gov  
Tissue Procurement: Gilbert Smith, Ph.D.  
CDNA Library Preparation: Life Technologies, Inc.  
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)  
DNA Sequencing by: Incyte Genomics, Inc.  
Clone distribution: MGC clone distribution information can be  
found through the I.M.A.G.E. Consortium/LIML at:  
<http://image.llnl.gov>  
Plate: LHAM8543 row: 1 column: 23  
High quality sequence stop: 724.  
FEATURES  
source Location/Qualifiers  
1..912  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="Czech II"  
/db\_xref="taxon:10090"  
/clone="IMAGE:3494566"  
/tissue\_type="spontaneous tumor, metastatic to mammary."  
Stem cell origin."  
/lab\_host="DH10B"  
/clone\_1ib="NCI CGAP Lu29"  
/note="Organ: lung; Vector: pCMV-Sport6; Site\_1: SalI;  
Site\_2: NotI; Cloned unidirectionally. Primer: Oligo dT.  
Library constructed by Life Technologies. Investigator  
providing samples: Gilbert Smith, NIH"

## ORIGIN

Query Match 17.9%; Score 217.8; DB 2; Length 912;  
Best Local Similarity 99.1%; Pred. No. 9,4e-44;  
Matches 219; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
Qy 1000 GTGAGCAGGAGCTGCTCTCTCTCAGACAGCTCCAGTGTGCCGGGAGAGAGAGA 1059  
Db 1 GTGAGCAGGAGCTGCTCTCTCTCAGACAGCTCCAGTGTGCCGGGAGAGAGAGA 60

Qy 1060 GAAAGACAGGCACTGGGAAAGAGCTGCTGCGGGACGGAGAAAGCTTCACTGATGAGAC 1119  
Db 61 GAAAGACAGGCACTGGGAAAGAGCTGCTGCGGGACGGAGAAAGCTTCACTGATGAGAC 120  
Qy 1120 TTATTACACGGGACAGCCCTGTGCTTATGACAGCAGCTGAGAGCTCAGGAGCAAGTTT 1179  
Db 121 TTATTACACGGGACAGCCCTGTGCTTATGACAGCAGCTGAGAGCTCAGGAGCAAGTTT 180  
Qy 1180 GCTGAACCTCAGAGTTTGAACCCAAAAGAGAGAGAAATG 1220  
Db 181 GCTGAACCTCAGAGTTTGAACCCAAAAGAGAGAGAAATG 221

RESULT 9  
BB569161 276 bp mRNA linear EST 29-NOV-2000  
LOCUS BB569161 RIKEN full-length enriched, 17 days embryo head Mus  
DEFINITION musculus cDNA clone 3322402G22 5', mRNA sequence.  
ACCESSION BB569161  
VERSION BB569161.1 GI:11460069  
KEYWORDS EST.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Murioidea; Muridae; Murinae; Mus.  
1 (bases 1 to 276)  
AUTHORS Aizawa, K., Akahira, S., Akimura, T., Arai, A., Arakawa, T.,  
Carninci, P., Hangaki, T., Hayatsu, N., Hiraoka, T., Hirozane, T.,  
Hodoyama, Y., Imotani, K., Ishii, Y., Itoh, M., Izawa, M., Kawai, D.,  
Kojima, Y., Konno, H., Kusakabe, M., Matsuyama, T., Miyazaki, A.,  
Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Okazaki, Y.,  
Okido, T., Owa, C., Sakai, C., Sakai, K., Sasaki, D., Sato, K.,  
Shibata, K., Shibata, Y., Shinagawa, A., Shiraki, T., Sogabe, Y.,  
Suzuki, H., Tagawa, A., Takahashi, F., Tanaka, T., Toya, T.,  
Watanabe, H., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshiki, A.,  
Muramatsu, M., and Hayashizaki, Y.

TITLE RIKEN Mouse ESTs (Aizawa, K. et al. 2000)  
JOURNAL Unpublished (2000)  
COMMENT Contact: Yoshinori Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216

Email: genome-research.riken.jp URL: <http://genome.gsc.riken.jp/>  
Carninci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoaka, S.,  
Sasaki, N., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.  
Thermolabile and thermolabile and thermolabile enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh, M., Kitanai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
Tomaru, Y., Carninci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,  
Okazaki, Y., and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carninci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (<http://genome.rtc.riken.go.jp>) for  
further details.  
FEATURES  
source Location/Qualifiers  
1..276  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="3322402G22"  
/sex="mixed"  
/tissue\_type="head"  
/dev\_stage="17 days embryo"  
/lab\_host="DH10B"



|||||  
Db 121 CCTGTGCTTAGACAGCAGCTAGAGACTGAGACCGCAAGTTTGCTGAAGTCTGACAGTTTA 180  
QY 1197 GAAACCCAAAAGAGAGAGAGATG 1220  
|||||  
Db 181 GAAACCCAAAAGAGAGAGAGATG 204  
|||||

RESULT 11  
AZ984331 661 bp DNA linear GSS 27-APR-2001  
AZ984331  
LOCUS 2M0265M18R Mouse 10kb plasmid UUGC2M library Mus musculus genomic  
DEFINITION clone UUGC2M0265M18 R, genomic survey sequence.  
ACCESSION AZ984331  
VERSION AZ984331.1 GI:13855558  
KEYWORDS GSS.  
SOURCE Mus musculus (house mouse)  
ORGANISM Mus musculus  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia;  
Sciurognathi; Muridae; Murinae; Mus.  
REFERENCE 1 (bases 1 to 661)  
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,  
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T.,  
Reilly,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von  
Niederhausern,A. and Wright,D., Weiss,R.  
Mouse whole genome scaffolding with paired end reads from 10kb  
Plasmid inserts  
Unpublished (2000)  
CONTACT: Robert B. Weiss  
University of Utah Genome Center  
University of Utah  
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLIC, UT  
84112, USA  
Tel: 801 585 5606  
Fax: 801 585 7177  
Email: ddunn@genetics.utah.edu  
Insert Length: 10000 Std Error: 0.00  
Plate: 0265 row: M column: 18  
Seq primer: CACACAGAAACAGCTATGACC  
Class: Plasmid ends  
High quality sequence stop: 661.  
Location/Qualifiers  
1..661  
/organism="Mus musculus"  
/mol\_type="genomic DNA"  
/strain="C57BL/6J"  
/db\_xref="taxon:10090"  
/clone="UUGC2M0265M18"  
/sex="Female"  
/lab\_host="E. coli strain XL10-Gold, T1-resistant, F-"  
/clone\_lib="Mouse 10kb plasmid UUGC2M library"  
/note="Vector: PWD42nv. Purified genomic DNA from M.  
musculus C57BL/6J (female) was obtained from the Jackson  
Laboratory Mouse DNA Resource  
(http://www.jax.org/resources/documents/dnares/). The DNA  
was hydrodynamically sheared by repeated passage through a  
0.005 inch orifice at constant velocity. The sheared DNA  
was blunt end-repaired with T4 DNA polymerase and T4  
polynucleotide kinase. Adaptor oligonucleotides were  
ligated to the blunt ends in high molar excess. The  
adapted DNA was purified and size-selected for a 9.5 to  
10.5 kb range using preparative agarose gel  
electrophoresis. Vector DNA was prepared from a derivative  
of PWD42 (G14732114|db|AF129072.1), a copy-number  
inducible derivative of plasmid R1. The vector was ligated  
with adaptors complementary to the insert adaptors and  
purified. The sheared, adapted mouse DNA was annealed to  
chemically-competent E. coli XL10-Gold (Stratagene) cells  
and selected for ampicillin resistance."

Query Match 16.7%; Score 203.4; DB 9; Length 661;  
Best Local Similarity 99.1%; Pred. No. 3,9e-40;  
Matches 215; Conservative 0; Mismatches 1; Indels 1; Gaps 1;  
QY 1004 GCAGGCTGCTCTCTCTCTCTCAGAGACAGCTCCGAGTGTCCCGGAGAAAGAGAAAG 1063  
|||||  
Db 8 GCCGGCTGCTCTCTCTCTCTCA-GACAGCTCCGAGTGTCCCGGAGAAAGAGAAAG 66  
|||||

QY 1064 AGACAGGCACTGGGAAAAGACCTGTGCGGAGCGGAAAGCTCTGACTGATGACTTAT 1123  
|||||  
Db 67 AGACAGGCACTGGGAAAAGACCTGTGCGGAGCGGAAAGCTCTGACTGATGACTTAT 126  
|||||

QY 1124 TCACAGGCAAGCCCTGTGCTCTTACACAGAGCTTACAGACCGCAAGTTTGCTG 1183  
|||||  
Db 127 TCACAGGCAAGCCCTGTGCTCTTACACAGAGCTTACAGACCGCAAGTTTGCTG 186  
|||||

QY 1184 AACTCAGTTTGAACCCAAAAGAGAGAGATG 1220  
|||||  
Db 187 AACTCAGTTTGAACCCAAAAGAGAGAGATG 223  
|||||

RESULT 12  
CR905089 793 bp DNA linear GSS 23-NOV-2004  
LOCUS CR905089  
DEFINITION Sus scrofa BES, genomic survey sequence.  
ACCESSION CR905089  
VERSION CR905089.1 GI:56229586  
KEYWORDS GSS; Bac-end sequence BES; Genome Survey Sequence.  
SOURCE Sus scrofa (pig)  
ORGANISM Sus scrofa  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Laurasiatheria; Cetartiodactyla; Suina; Suidae;  
Sus.

REFERENCE 1 (bases 1 to 793)  
Rogel-Galliard,C., Bourdeaux,N., Billaule,A., Vaiman,M. and  
Chardon,P.  
Construction of a swine BAC library: application to the  
characterization and mapping of porcine type C endoviral elements  
Cytogenet. Cell Genet. 85 (3-4), 205-211 (1999)  
10449899  
PUBMED  
JOURNAL  
AUTHORS 2 (bases 1 to 793)  
Chardon,P., Iannuccelli,N., Roig,A., Dosset,C., Demars,J.,  
Rogel-Galliard,C., Roy,A., Schibler,L. and Milan,D.  
A physical map of the swine genome  
Unpublished  
3 (bases 1 to 793)  
TITLE Direct Submission  
JOURNAL Submitted (18-NOV-2004) Genoscope - Centre National de Sequencage :  
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr  
- Web : www.genoscope.cns.fr)  
JOURNAL  
AUTHORS  
TITLE  
JOURNAL  
AUTHORS

FEATURES  
source 1..793  
Location/Qualifiers  
1..793  
/organism="Sus scrofa"  
/mol\_type="genomic DNA"  
/strain="Large White"  
/db\_xref="taxon:9823"  
/clone="B10352F04"  
/sex="male"  
/cell\_type="fibroblast"  
/clone\_lib="SBAB"  
/note="Genoscope sequence ID : IH0AA43DD10FM1"

ORIGIN  
Query Match 15.2%; Score 185.8; DB 11; Length 793;  
Best Local Similarity 65.5%; Pred. No. 1.1e-35;  
Matches 476; Conservative 0; Mismatches 202; Indels 49; Gaps 12;  
QY 290 TATGTATTTGTTACTTAACAGACATTAATCAGGCCAAGCACTTAAAGTTTATTTGAT 349  
|||||  
Db 51 TATGTATTTAAATTATGAGGGCCATTATCAATGTCAGCAATTTTAAATTACTTGAAT 110  
|||||  
QY 350 A-----GTATTTGTGTAAAGGCAAGACATGTGATGCTGAGAAAAAAGTATTTGGT 402  
|||||

```

Db      111 ACTGAAGTGGATATTTTGTGTTAAAGATTTAAAGTGAAGCTCAAAACCTGATTTTCCT 170
Qy      403 AACTGATTAAATA-----TATCAACTGGGTTAAATTAATAAAAAA----- 445
Db      171 AACCCATTTTAATAAGTAGATGTTTCTATCCAAATGTTTAAAGTAGAAAAATTCGCAT 230
Qy      446 ---TGTCATTAAGTAAAAAACCATAATACCAACAAGACTTAATCTCCCTGGAAA 502
Db      231 TCTTAAGCTAGTGGTTCAAAATAACAGTAATACCAAGAACTATTTCCCTATGAAAA 290
Qy      503 AGCACA-----TTTACAAGGGCTGATCTTAGCCCTTATATTTACATTAAGAAATAAAC 558
Db      291 GGCAGAAAGATTTTAAGAGACAGACTTCTTCCTTTTGTGTAATACAAATAAAC 350
Qy      559 AAGGTCGGATATAGCTGTATTTTATTCCT-AAAAGAACAGAACTTCACTATGCTT 617
Db      351 AAGGTCCTGACATATTTGATGATTTTATTCCTTAAGAAAAACAGAACTTTCATTG 410
Qy      618 AAAATTAAGTATTAAGTCTGATGATCTGCAAGCTTACCC--TACAAACGACAGACA 674
Db      411 TACTTAAAGTATGCCCCAGATTTTACAGCTCAGCCCTGTGAAATCAGTTTCAGA 470
Qy      675 GACAAAGAGCCCAAGCTACTCTCTAGAAAT-----ATTAGGGTGGCTCTGACAG 730
Db      471 CACAAAGACCCGCAAGCTCTCTAGAAATCTCAATTAAGGAGTGTCTCTGAAATA 530
Qy      731 CCCAGGGCTCTGTGGCTGGCTG- TGTTCCAGAAAGGCTTGTGACAGAGTCT- T 788
Db      531 CCCAGGGGCTGTGAATCATGATGAGGCTTTTCCAGAGGGCTTCTACAGCTATAGCCCTG 590
Qy      789 TGGGACAGTAAGCATATGCTGATTTTCTGTGCTGCTGCTAGTGAACCCCTACAG 848
Db      591 CTGGACACACACACACACACCTGGAATTTTCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 650
Qy      849 AAGATAGTGGTGGAGCAGG-----GGGCGAGCGGCTGCTGACATGCTGGCTGT 902
Db      651 AAGATAGGAGCAAGCGCGGAGGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 709
Qy      903 CTTATCAACTTATCATATTAAGGAAAGAAAGTATGATGATGATGATGATGATGATGATG 962
Db      710 CTTATCAACTTATCATATTA- GAAAAAGAAAGTATGATGATGATGATGATGATGATGATG 768
Qy      963 TCAGGGG 969
Db      769 TCTGGGG 775

```

**RESULT 13**  
**LOCUS** BX506958  
**DEFINITION** DKFZp779M112\_r1 779 (synonym: hnccl) Homo sapiens cDNA clone  
**ACCESSION** BX506958  
**VERSION** BX506958.1 GI:32047724  
**KEYWORDS** EST.  
**SOURCE** Homo sapiens (human)  
**ORGANISM** Homo sapiens  
**REFERENCE** EST (Poustka, A., Albert, R., Moosmayer, P., Schupp, I., Wellenreuther, R., Wiemann, S., Poustka, A., Albert, R., Moosmayer, P., Schupp, I., Wellenreuther, R., Meues, H.W., Well, B., Amid, C., Osanger, A., Fobo, G., Han, M. and Wiemann, S.)  
**AUTHORS** Ingolettaeder Landstr. 1, D-85764 Neuherberg, Germany  
**JOURNAL** This is the 5' sequence of the clone insert  
**COMMENT** Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

**FEATURES**  
**source**  
 Research Center (DKFZ), Email s.wiemann@dkfz-heidelberg.de; sequenced by DKFZ (German Cancer Research Center, Heidelberg/Germany) within the cDNA sequencing consortium of the German Genome Project.  
 No 5' sequence available.  
 This clone (DKFZp779M112) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.  
 location/Qualifiers  
 1..726  
 /organism="Homo sapiens"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:9606"  
 /clone="DKFZp779M112"  
 /tissue\_type="liver"  
 /dev\_stage="fetal"  
 /lab\_host="DH10B"  
 /clone\_id="779 (synonym: hnccl)"  
 /note="Vector: pSport1\_Sfi; Site\_1: SfiI; Site\_2: SfiIB"

**ORIGIN**  
 Query Match  
 Best Local Similarity 75.4%; Pred. No. 7,4e-34;  
 Matches 248; Conservative 0; Mismatches 79; Indels 2; Gaps 2;

```

Qy      706 TAATTAAGGCTGCTGCTGACATGCCCCAGGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTG 764
Db      46 TAATGAGGCTGCTGCTGCTGAGAAATGCCAGGGCTCTGTAACATCGGTTTTCCTCCAGA 105
Qy      765 AGGCTTTCAGAGTACACAGTCTTTGGGAGTAAAGCATATGCTCTGATTTTTCCTTT 824
Db      106 GGGTTTTCAGACATGAGGCTCTGTTGAGGCGAGGCAATTTGCTGATTTTTCCTGTT 165
Qy      825 GCTGCTAGTGAACCCCTTACAGAAAGATAGTGGTGGAGCAGG- GGGCGAGCGGCTGG 883
Db      166 GCTGCTAGTGAACCCCTTACAGAAAGATAGTGGTGGAGCAGG- GGGCGAGCGGCTGG 225
Qy      884 CTGACATGCTGCTGCTCTTATCAACTTATCATATTAAGGAAAGAAAGTATGATGATG 943
Db      226 CTACACATGCTGCTGCTCTTATCAACTTATCATATTAAGGAAAGAAAGTATGATGATG 285
Qy      944 GGATTCAGACCTGATGAGCTGAGGGGAGAAACAAGAGTCCGTGCACAGCTCTGAGTGA 1003
Db      286 GGATTCAGACCTGATGAGCTGAGGGGAGAAACAAGAGTCCGTGCACAGCTCTGCTGT 345
Qy      1004 GCAGGCTGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 345
Db      346 AAAAGCTGACACAGCCCTCCCAAGTGAGC 374

```

**RESULT 14**  
**LOCUS** BB601394  
**DEFINITION** BB601394 RIKEN full-length enriched, 13 days embryo lung Mus musculus cDNA clone D43011106 5', mRNA sequence.  
**ACCESSION** BB601394  
**VERSION** BB601394.1 GI:11509995  
**KEYWORDS** EST.  
**SOURCE** Mus musculus (house mouse)  
**ORGANISM** Mus musculus  
**REFERENCE** Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Euarchontoglires; Glires; Rodentia; Sciurognathi; Muridae; Muridae; Murinae; Mus.  
 1 (bases 1 to 204)  
 Alizawa, K., Akahira, S., Akimura, T., Arai, A., Arakawa, T., Carninci, P., Haraoka, T., Hayatsu, N., Hirooka, T., Hirozane, T., Hodoiyama, Y., Imotani, K., Ishii, Y., Itoh, M., Itawa, M., Kawai, J., Kojima, Y., Kono, H., Kusakabe, M., Matsuyama, T., Miyazaki, A., Nakamura, M., Nishi, K., Nomura, K., Numazaki, R., Okazaki, Y., Okido, T., Owa, C., Sakai, C., Sakai, K., Sasaki, D., Sato, K., Shibata, K., Shibata, Y., Shingawa, A., Shiraki, T., Sogabe, Y., Suzuki, H., Tagawa, A., Takehashi, F., Tanaka, T., Toya, T., Watabiki, A., Yamamura, T., Yasunishi, A., Yoshida, K., Yoshiki, A.,

TITLE  
JOURNAL  
COMMENT

Muramatsu, M. and Hayashizaki, Y.  
RIKEN Mouse ESTs (Aizawa, K. et al. 2000)  
Unpublished (2000)  
Contact: Yoshihide Hayashizaki  
Laboratory for Genome Exploration Research Group, RIKEN Genomic  
Sciences Center (GSC), Yokohama Institute  
The Institute of Physical and Chemical Research (RIKEN)  
1-7-22 Suenho-cho, Tsurumi-Ku, Yokohama, Kanagawa 230-0045, Japan  
Tel: 81-45-503-9222  
Fax: 81-45-503-9216  
Email: genome-res@sc.riken.jp, URL: http://genome-gsc.riken.jp/  
Carinci, P., Nishiyama, Y., Westover, A., Itoh, M., Nagaoaka, S.,  
Sasaki, N., Okazaki, Y., Muramatsu, M. and Hayashizaki, Y.  
Thermolabile and thermolabile activation of thermolabile enzymes by  
trehalose and its application for the synthesis of full length  
cDNA. Proc. Natl. Acad. Sci. U.S.A. 95 (2), 520-524 (1998)  
Itoh, M., Kitsuai, T., Akiyama, J., Shibata, K., Izawa, M., Kawai, J.,  
Tomaru, Y., Carinci, P., Shibata, Y., Ozawa, Y., Muramatsu, M.,  
Okazaki, Y. and Hayashizaki, Y.  
Automated filtration-based high-throughput plasmid preparation  
system. Genome Res. 9 (5), 463-470 (1999)  
Carinci, P. and Hayashizaki, Y.  
High-efficiency full-length cDNA cloning. Methods Enzymol. 303,  
19-44 (1999)  
Please visit our web site (http://genome.rtc.riken.go.jp) for  
further details.

FEATURES  
source

Location/Qualifiers  
1..204  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/db\_xref="taxon:10090"  
/clone="D430011v06"  
/tissue\_type="lung"  
/dev\_stage="13 days embryo"  
/lab\_host="DH10B"  
/clone\_1lb="RIKEN full-length enriched, 13 days embryo  
lung"  
/note="Site\_1: SalI; Site\_2: BamHI; cDNA library was  
prepared and sequenced in Mouse Genome Encyclopedia  
Project of Genome Exploration Research Group in Riken  
Genomic Sciences Center and Genome Science Laboratory in  
RIKEN. Division of Experimental Animal Research in Riken  
contributed to prepare mouse tissues. 1st strand cDNA was  
primed with a primer 15'.  
GAGAGAGAGAGCGCGCACTGAGTCTTTTCTTTTCTTTT 3'. cDNA was  
transcribed by using trehalose thermo-activated reverse  
transcriptase and subsequently enriched for full-length by  
cap-trapper. Second strand cDNA was prepared with the  
primer adapter of sequence 15'.  
GAGAGAGAGATTCGAGTTAAATTAATTAATCCCCCCCC 3'. cDNA  
was cleaved with BamHI and XhoI. Vector: a modified  
pBluescript KS(+) after bulk excision from Lambda P1C I."

ORIGIN

Query Match 14.4%; Score 175.8; DB 2; Length 204;  
Best Local Similarity 91.6%; Pred. No. 2.9e-33;  
Matches 186; Conservative 0; Mismatches 17; Indels 0; Gaps 0;  
948 ACTGACACTGTAGCTCAGGAGGAAAGAGTCCGTCAGACTCTGAGTGGAGC 1007  
|||||  
2 ACTGACACTGTAAATCAGGGAGAAACAAAGATCCGTCAGAACTCTGAGTGGAGC 61  
|||||  
1008 GGCCTCTCTCTCTCTCAGAGACGCTCCGAGTGTGCGCGGAGAAAGAGAGAGAC 1067  
|||||  
62 GGGTGTCTCTCTCTCAGAGAAAGCTCCGAGTGTGCGCGGAGAAAGAGAGAGAC 121  
|||||  
1068 AGGCACTGGGAAGAGCCCTGCTCGGAGAGGAGGAGGCTCTCATGATGACTTATTCAC 1127  
|||||  
122 AGGGAAGTGGGAAGAGCCCTGCTCGGAGAGGAGGAGGAGGCTCATGATGACTTATTCAC 181  
|||||  
1128 AGGCAAGAGCCCTGCTGAGCTTAGA 1150  
|||||  
DB 182 ACGGAGCAGAGCCCTGTGGCTCAGA 204  
|||||

RESULT 15  
BUS80281 525 bp mRNA linear EST 17-SEP-2002  
LOCUS  
DEFINITION  
BK7909.y1 Kaestner ngn3 wt Mus musculus cDNA 5' similar to  
IR:035608 035608 ANGIOPOIETIN 2.; mRNA sequence.  
ACCESSION  
BUS80281  
VERSION  
BUS80281.1 GI:23065508  
KEYWORDS  
SOURCE  
EST.  
ORGANISM  
Mus musculus (house mouse)

REFERENCE  
AUTHORS

Melton, D., Brown, J., Kently, G., Permutt, A., Lee, C., Kaestner, K.,  
Lemishka, I., Scaer, M., Brestelli, J., Gradwohl, G., Clifton, S.,  
Hillier, L., Maira, M., Page, D., Wylie, T., Martin, J., Blistain, A.,  
Schmitt, A., Theising, B., Ritzer, E., Ronko, I., Bennett, J.,  
Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagaris, R.,  
Williams, T., Jackson, Y. and Bowers, Y.  
Endocrine Pancreas Consortium  
Unpublished (2000)

TITLE

JOURNAL

COMMENT

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue  
Endocrine Pancreas Consortium  
Harvard University, Howard Hughes Medical Institute  
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,  
MA 02138  
Tel: 617-495-1812  
Fax: 617-495-8557  
Email: dmelton@biohp.harvard.edu  
Pancreas was obtained from Gerard Gradwohl (PNAS 97 P1607-1611,  
2000) Library was constructed by Catherine Lee DNA sequencing by:  
Washington University Genome Sequencing Center For information on  
obtaining a clone please contact: Dr. Marie Scaer  
(mscaer@mail.med.upenn.edu)  
Seq primer: -40RP from Gibco  
High quality sequence stop: 343.  
Location/Qualifiers  
1..525  
/organism="Mus musculus"  
/mol\_type="mRNA"  
/strain="129/Sv x CD1"  
/db\_xref="taxon:10090"  
/dev\_stage="P.c. 14.5"  
/lab\_host="E. coli-DH12S (GIBCO)"  
/clone\_1lb="Kaestner ngn3 wt"  
/note="Organ: pancreas; Vector: pSPORT1 (GIBCO); Site\_1:  
Not I; Site\_2: Sal I; The library was prepared by  
Catherine S. Lee and has not been published. The pancreas  
was obtained from Gerard Gradwohl (PNAS 97 P1607-1611,  
2000). The cDNA's were prepared with an oligo containing a  
NotI site, and SalI linkers were added to the ends. The  
inserts were cut with NotI before being cloned into the  
NotI-SalI sites in the vectors. This is one of two  
libraries, ngn3 wt and ngn3 -/- . The wt library is in  
pSPORT1, T7 promoter is 5'."

FEATURES  
source

ORIGIN

Query Match 13.8%; Score 167.8; DB 5; Length 525;  
Best Local Similarity 96.1%; Pred. No. 3.7e-31;  
Matches 172; Conservative 0; Mismatches 7; Indels 0; Gaps 0;  
1042 GCGGGGAG 1101  
|||||  
35 GCGAGTACGGGGAG 94  
|||||  
1102 AGGCTCTACGATGAGCTTATTTCACAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 1161  
|||||  
DB 95 AGGCTCTACGATGAGCTTATTTCACAGGAGAGAGAGAGAGAGAGAGAGAGAGAG 154  
|||||  
QY 1162 AGCTGAG 1220  
|||||

Db 155 AGCTCAGGAGGCAAGTTGCTGAACTCAGTTTGAACCCAAAAGAGAGAGATG 213

Search completed: December 12, 2005, 15:19:28  
Job time : 4730 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 12, 2005, 12:00:02 ; Search time 260 Seconds  
(without alignments)  
8340.866 Million cell updates/sec

Title: US-10-627-075-1

Perfect score: 1220  
Sequence: 1 gagcttccagctaccgacgac.....ccaaaagagagagagatg 1220

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents NA: \*  
1: /cgn2\_6/prodata/1/ina/1/COMB.seq: \*  
2: /cgn2\_6/prodata/1/ina/5/COMB.seq: \*  
3: /cgn2\_6/prodata/1/ina/6A/COMB.seq: \*  
4: /cgn2\_6/prodata/1/ina/6B/COMB.seq: \*  
5: /cgn2\_6/prodata/1/ina/H/COMB.seq: \*  
6: /cgn2\_6/prodata/1/ina/PC/COMB.seq: \*  
7: /cgn2\_6/prodata/1/ina/PP/COMB.seq: \*  
8: /cgn2\_6/prodata/1/ina/RE/COMB.seq: \*  
9: /cgn2\_6/prodata/1/ina/backfile1.seq: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	47.4	3.9	7218	2	US-08-232-463-14 Sequence 14, Appl
C 2	46.4	3.8	1864	3	US-09-468-265-4 Sequence 4, Appl
C 3	46.2	3.8	1141	3	US-09-806-708B-22 Sequence 22, Appl
C 4	45.4	3.7	20674	3	US-09-641-638-651 Sequence 651, App
C 5	45.4	3.7	20674	3	US-10-170-097-651 Sequence 651, App
C 6	44.6	3.7	732	3	US-08-998-416-1036 Sequence 1036, App
C 7	44	3.6	1141	3	US-09-806-708B-22 Sequence 22, Appl
C 8	43.8	3.6	2308	3	US-10-104-047-682 Sequence 682, App
C 9	42.6	3.5	80411	3	US-09-949-016-15777 Sequence 15777, A
C 10	42.4	3.5	3493	3	US-09-949-016-12063 Sequence 12063, A
C 11	42.4	3.5	95020	3	US-09-949-016-13272 Sequence 13272, A
C 12	42.2	3.5	50000	3	US-09-662-254B-25 Sequence 25, Appl
C 13	42	3.4	601	3	US-09-949-016-83375 Sequence 83375, A
C 14	42	3.4	601	3	US-09-949-016-83376 Sequence 83376, A
C 15	42	3.4	143550	3	US-09-949-016-14143 Sequence 14143, A
C 16	41.8	3.4	832	3	US-09-621-976-2813 Sequence 2813, Ap
C 17	41.6	3.4	280	3	US-09-662-254B-27 Sequence 27, Appl
C 18	41.6	3.4	32392	3	US-09-662-254B-24 Sequence 24, Appl
C 19	41.2	3.4	50000	3	US-09-662-254B-24 Sequence 24, Appl
C 20	41	3.4	244	3	US-09-621-976-484 Sequence 484, Appl
C 21	40.6	3.3	19124	2	US-08-487-826B-13 Sequence 13, Appl
C 22	40.6	3.3	601	3	US-09-949-016-70227 Sequence 70227, A
C 23	40.6	3.3	70323	3	US-09-949-016-17594 Sequence 17594, A
C 24	40.4	3.3	837	3	US-08-998-416-288 Sequence 288, App

C 25	40.4	3.3	1412	3	US-09-270-767-7491 Sequence 7491, Ap
C 26	40.4	3.3	1412	3	US-09-270-767-22773 Sequence 22773, A
C 27	40.4	3.3	6996	3	US-09-573-080A-29 Sequence 29, Appl
C 28	40.4	3.3	278866	3	US-09-949-016-13922 Sequence 13922, A
C 29	40.4	3.3	278866	3	US-09-949-016-13923 Sequence 13923, A
C 30	40.4	3.3	278866	3	US-09-949-016-13924 Sequence 13924, A
C 31	40.4	3.3	278866	3	US-09-949-016-13925 Sequence 13925, A
C 32	40.4	3.3	278866	3	US-09-949-016-13926 Sequence 13926, A
C 33	40.4	3.3	278866	3	US-09-949-016-14699 Sequence 14699, A
C 34	40.4	3.3	278866	3	US-09-949-016-14700 Sequence 14700, A
C 35	40.4	3.3	278866	3	US-09-949-016-14701 Sequence 14701, A
C 36	40.4	3.3	278866	3	US-09-949-016-14702 Sequence 14702, A
C 37	40.4	3.3	278866	3	US-09-949-016-14703 Sequence 14703, A
C 38	40.2	3.3	601	3	US-09-949-016-206713 Sequence 206713, Appl
C 39	40.2	3.3	11748	2	US-08-611-107-30 Sequence 30, Appl
C 40	40	3.3	663	3	US-08-998-416-191 Sequence 191, App
C 41	40	3.3	921	3	US-09-248-796A-9583 Sequence 9583, Ap
C 42	40	3.3	7286	3	US-09-331-581-13 Sequence 3, Appl
C 43	40	3.3	7938	3	US-09-331-581-14 Sequence 14, Appl
C 44	40	3.3	72278	3	US-09-949-016-16113 Sequence 16113, A
C 45	39.8	3.3	19124	2	US-08-487-826B-13 Sequence 13, Appl

#### ALIGNMENTS

RESULT 1  
US-08-232-463-14/c  
; Sequence 14, Application US/08232463  
; Patent No. 5670367  
; GENERAL INFORMATION:  
; APPLICANT: DORNER, F.  
; APPLICANT: SCHIEFLINGER, F.  
; APPLICANT: FALKNER, F. G.  
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS  
; NUMBER OF SEQUENCES: 52  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 1800 Diagonal Road, Suite 500  
; CITY: Alexandria  
; STATE: VA  
; COUNTRY: USA  
; ZIP: 22313-0299  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/232,463  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US/07/935,313  
; FILING DATE:  
; APPLICATION NUMBER: EP 91 114 300.6  
; FILING DATE: 26-AUG-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: BENT, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (703) 836-9300  
; TELEFAX: (703) 683-4109  
; TELEX: 899149  
; INFORMATION FOR SEQ ID NO: 14:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 7218 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; IMMEDIATE SOURCE:  
; CLONE: PTZgpc-PIs







NAME/KEY: allele  
LOCATION: 3802  
OTHER INFORMATION: 12-206-81 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 4062  
OTHER INFORMATION: 10-343-231 : deletion of C  
NAME/KEY: allele  
LOCATION: 4088  
OTHER INFORMATION: 12-206-366 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 4109  
OTHER INFORMATION: 10-343-278 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 4170  
OTHER INFORMATION: 10-343-339 : polymorphic base G or T  
NAME/KEY: allele  
LOCATION: 5903  
OTHER INFORMATION: 10-346-23 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 6019  
OTHER INFORMATION: 10-346-141 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 6141  
OTHER INFORMATION: 10-346-263 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 6183  
OTHER INFORMATION: 10-346-305 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 6338  
OTHER INFORMATION: 10-347-74 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 6375  
OTHER INFORMATION: 10-347-111 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 6429  
OTHER INFORMATION: 10-347-165 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 6467  
OTHER INFORMATION: 10-347-203 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 6484  
OTHER INFORMATION: 10-347-220 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 6534  
OTHER INFORMATION: 10-347-271 : polymorphic base A or T  
NAME/KEY: allele  
LOCATION: 6611  
OTHER INFORMATION: 10-347-348 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 7668  
OTHER INFORMATION: 10-348-391 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 8608  
OTHER INFORMATION: 10-349-47 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 8658  
OTHER INFORMATION: 10-349-97 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 8703  
OTHER INFORMATION: 10-349-142 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 8777  
OTHER INFORMATION: 10-349-216 : deletion of CTG  
NAME/KEY: allele  
LOCATION: 8785  
OTHER INFORMATION: 10-349-224 : polymorphic base G or T  
NAME/KEY: allele  
LOCATION: 8926  
OTHER INFORMATION: 10-349-368 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 12171  
OTHER INFORMATION: 10-350-72 : polymorphic base C or T  
NAME/KEY: allele

LOCATION: 12429  
OTHER INFORMATION: 10-350-332 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 13341  
OTHER INFORMATION: 10-507-170 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 13492  
OTHER INFORMATION: 10-507-321 : polymorphic base A or C  
NAME/KEY: allele  
LOCATION: 13524  
OTHER INFORMATION: 10-507-353 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 13535

Query Match  
Best Local Similarity 49.0%; Pred. No. 0.038;  
Matches 180; Conservative 0; Mismatches 181; Indels 6; Gaps 2;

QY 276 ATACTGACGCAATATCTATGTTACTTACAGACATTTATCAGGCCAAACACTTTA 335  
DB 11439 AAAATTTAATTAATTAATGAAGCAATTAATTAATTAATTAATTAATTAATTA 11380

QY 336 AGTTTATTTGATATGATTTTGTGTTAAGCACAAGATGAGTCTGAGAAACTGA 395  
DB 11379 AATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11320

QY 396 TGTGTGTAACCTGATTTAATA-ATATCAACGTGGTTAAATTAATTAATTAATTAATTA 453  
DB 11319 TTAATTAAGTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11260

QY 454 ACTTAAAAAACAACCAATTCACAAAGCTT---ACTGCCCTGGAAAAAGCAGAT 509  
DB 11259 AATTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 11200

QY 510 TTACAAGGCTGATCTTACGCTTATATTTTACAATTAAGAAATTAACCAAGTCCGAT 569  
DB 11199 TAAATTTATTAATTTTAAATTTATTTTAAATTAATTAATTAATTAATTAATTAAT 11140

QY 570 ATAGCTGAATTTATTCCTTAAAGAACAGAACTTCACTATGCTTTAAATTAAGTG 629  
DB 11139 AAAATTTTAAATTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTTTC 11080

QY 630 ATTACCT 636  
DB 11079 CTTAGCT 11073

RESULT 5  
US-10-170-097-651/C  
Sequence 651, Application US/10170097  
Patent No. 6794143  
GENERAL INFORMATION:  
APPLICANT: Blumenfeld, Marta  
APPLICANT: Bougueleret, Lydie  
APPLICANT: Chumakov, Ilya  
APPLICANT: Cohen, Annick  
TITLE OR INVENTION: BIALLERIC MARKERS DERIVED FROM GENOMIC REGIONS CARRYING  
FILE REFERENCE: GEN-T114XC2D1  
CURRENT APPLICATION NUMBER: US/10/170,097  
PRIOR FILING DATE: 2002-06-10  
PRIOR APPLICATION NUMBER: US 09/641,638  
PRIOR FILING DATE: 2000-08-16  
PRIOR APPLICATION NUMBER: US 09/502,330  
PRIOR FILING DATE: 2000-02-11  
PRIOR APPLICATION NUMBER: US 60/133,200  
PRIOR FILING DATE: 1999-05-07  
PRIOR APPLICATION NUMBER: US 09/275,267  
PRIOR FILING DATE: 1999-03-23  
PRIOR APPLICATION NUMBER: US 60/119,917  
PRIOR FILING DATE: 1998-02-12  
NUMBER OF SEQ ID NOS: 1304  
SOFTWARE: Patent.pm

```
; SEQ ID NO 651
;
; LENGTH: 20674
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 1123..3123
; OTHER INFORMATION: 5'regulatory region
; FEATURE:
; NAME/KEY: exon
; LOCATION: 3124..3297
; OTHER INFORMATION: exon 1
; FEATURE:
; NAME/KEY: exon
; LOCATION: 3871..4072
; OTHER INFORMATION: exon 2
; FEATURE:
; NAME/KEY: exon
; LOCATION: 5552..5633
; OTHER INFORMATION: exon 3
; FEATURE:
; NAME/KEY: exon
; LOCATION: 5758..5880
; OTHER INFORMATION: exon 4
; FEATURE:
; NAME/KEY: exon
; LOCATION: 5996..6099
; OTHER INFORMATION: exon 5
; FEATURE:
; NAME/KEY: exon
; LOCATION: 6349..6509
; OTHER INFORMATION: exon 6
; FEATURE:
; NAME/KEY: exon
; LOCATION: 7379..7522
; OTHER INFORMATION: exon 7
; FEATURE:
; NAME/KEY: exon
; LOCATION: 8645..8854
; OTHER INFORMATION: exon 8
; FEATURE:
; NAME/KEY: exon
; LOCATION: 12254..12340
; OTHER INFORMATION: exon 9
; FEATURE:
; NAME/KEY: exon
; LOCATION: 12854..13023
; OTHER INFORMATION: exon 10
; FEATURE:
; NAME/KEY: exon
; LOCATION: 13308..13429
; OTHER INFORMATION: exon 11
; FEATURE:
; NAME/KEY: exon
; LOCATION: 16567..16667
; OTHER INFORMATION: exon 12
; FEATURE:
; NAME/KEY: exon
; LOCATION: 16775..16945
; OTHER INFORMATION: exon 13
; FEATURE:
; NAME/KEY: exon
; LOCATION: 17063..17554
; OTHER INFORMATION: exon 14
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: 17555..20674
; OTHER INFORMATION: 3'regulatory region
; FEATURE:
; NAME/KEY: allele
; LOCATION: 1128
; OTHER INFORMATION: 10-508-191 : polymorphic base C or T
; FEATURE:
```

```
; NAME/KEY: allele
; LOCATION: 1182
; OTHER INFORMATION: 10-508-245 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 1559
; OTHER INFORMATION: 10-509-284 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 1570
; OTHER INFORMATION: 10-509-295 : deletion of C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 1827
; OTHER INFORMATION: 10-510-173 : variable motif ATTAA or TTTTTT
; FEATURE:
; NAME/KEY: allele
; LOCATION: 2048
; OTHER INFORMATION: 10-511-62 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 2323
; OTHER INFORMATION: 10-511-337 : insertion of T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 2341
; OTHER INFORMATION: 10-512-36 : polymorphic base G or C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 2623
; OTHER INFORMATION: 10-512-318 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 2832
; OTHER INFORMATION: 10-513-250 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 2844
; OTHER INFORMATION: 10-513-262 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 2934
; OTHER INFORMATION: 10-513-352 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 2947
; OTHER INFORMATION: 10-513-365 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 3802
; OTHER INFORMATION: 12-206-81 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
; LOCATION: 4062
; OTHER INFORMATION: 10-343-231 : deletion of C
; FEATURE:
; NAME/KEY: allele
; LOCATION: 4088
; OTHER INFORMATION: 12-206-366 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 4109
; OTHER INFORMATION: 10-343-278 : polymorphic base C or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 4170
; OTHER INFORMATION: 10-343-339 : polymorphic base G or T
; FEATURE:
; NAME/KEY: allele
; LOCATION: 5903
; OTHER INFORMATION: 10-346-23 : polymorphic base A or G
; FEATURE:
; NAME/KEY: allele
```

```
LOCATION: 6019
OTHER INFORMATION: 10-346-141 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 6141
OTHER INFORMATION: 10-346-263 : polymorphic base G or C
FEATURE:
NAME/KEY: allele
LOCATION: 6183
OTHER INFORMATION: 10-346-305 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 6338
OTHER INFORMATION: 10-347-74 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 6375
OTHER INFORMATION: 10-347-111 : polymorphic base G or C
FEATURE:
NAME/KEY: allele
LOCATION: 6429
OTHER INFORMATION: 10-347-165 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 6467
OTHER INFORMATION: 10-347-203 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 6484
OTHER INFORMATION: 10-347-220 : polymorphic base A or G
```

```
Query Match 3.7%; Score 45.4; DB 3; Length 20674;
Best Local Similarity 49.0%; Pred. No. 0.038;
Matches 180; Conservative 0; Mismatches 181; Indels 6; Gaps 2;
```

```
QY 276 ATATCTCAGCAATATCTATTGTTAATTAACAGACATTAATCAGCCAAACCACTTTA 335
DB 11439 AATATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 11380
QY 336 AGTTTATTTGTAATTTGTTAGTGTGTTAGGCAACATGAGTGTGAGAAACTGA 395
DB 11379 AATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 11320
QY 396 TGTGTGAATCTGATTTAATA--ATATCAAACTGGTGTAAATTAATTAATTAATTA 453
DB 11319 TTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 11260
QY 454 ACTTAAAAAACAATTAACCAAGACTTT---ACTTCCCTTGGAAAAACACAT 509
DB 11259 ATTATATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 11200
QY 510 TTACAAGGCGTATCTTACCTTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 569
DB 11199 TAAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 11140
QY 570 ATAGCTGTAATTTATCTTAAGAACAGAACTTCACTATGCTTTAAATTAAGTG 629
DB 11139 AAAAAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTT 629
QY 630 ATTACT 636
DB 11079 CTTAGCT 11073
```

```
RESULT 6
US-08-998-416-1036/c
; Sequence 1036, Application US/08998416
; Patent No. 6239264
```

```
; GENERAL INFORMATION:
; APPLICANT: Philippsen, Peter
; APPLICANT: Bohlmann, Rainer
; APPLICANT: Steiner, Sabine
; APPLICANT: Mohr, Christine
```

```
APPLICANT: Wendland, Jürgen
APPLICANT: Knechtle, Philipp
APPLICANT: Redischung, Corinne
TITLE OF INVENTION: GENOMIC DNA SEQUENCES OF ASHBYA GOSYPYII
TITLE OF INVENTION: AND USES THEREOF
NUMBER OF SEQUENCES: 1152
CORRESPONDENCE ADDRESS:
ADDRESSER: No. 6239264artis Corporation
STREET: 3054 Cornwallis Road
CITY: Research Triangle Park
STATE: No. 6239264th Carolina
COUNTRY: USA
ZIP: 27709
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/998,416
FILING DATE: 24-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: CH 0016/97
FILING DATE: 31-DEC-1996
ATTORNEY/AGENT INFORMATION:
NAME: Meigs, J. Timothy
REGISTRATION NUMBER: 38,241
REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
TELECOMMUNICATION INFORMATION:
TELEPHONE: 919-541-8587
TELEFAX: 919-541-8689
INFORMATION FOR SEQ ID NO: 1036:
SEQUENCE CHARACTERISTICS:
LENGTH: 732 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORIGINAL SOURCE:
ORGANISM: PAG1633UP
US-08-998-416-1036
```

```
Query Match 3.7%; Score 44.6; DB 3; Length 732;
Best Local Similarity 46.3%; Pred. No. 0.01;
Matches 146; Conservative 0; Mismatches 169; Indels 0; Gaps 0;
```

```
QY 339 TTATTTGTAATTTGTTAGTGTGTTAGGCAACATGAGTGTGAGAAACGTGAT 398
DB 732 TCTATATTAATTAATTTGTTATCTAATTAATTAATTAATTAATTAATTAATTA 673
QY 399 TGTGAATCTGATTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 458
DB 672 TTTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 613
QY 459 AAAAAAACAATTAACCAAGACTTTACTTCCCTTGGAAAAACACATTTACAAGG 518
DB 612 CTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 553
QY 519 CTGATCTTACCTTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 578
DB 552 AATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 493
QY 579 ATTATTTCTTAAGAACAGAACTTCACTATGCTTTAAATTAAGTGAATTA 638
DB 492 AGTAATATCAATTTAAGGTAAGTGAATTAATTAATTAATTAATTAATTAATTA 433
QY 639 GATACCTGCAAGCT 653
DB 432 AATATTAATTAATTAATTTAATTTAATTTAATTTAATTTAATTTAATTTAATTTA 418
```

```
RESULT 7
```

```
US-09-806-708B-22
; Sequence 22: Application US/09806708B
; Patent No. 6784342
; GENERAL INFORMATION:
; APPLICANT: The University of British Columbia
; TITLE OF INVENTION: Regulation of Embryonic Transcription in Plants
; FILE REFERENCE: 4810-58741
; CURRENT APPLICATION NUMBER: US/09/806,708B
; CURRENT FILING DATE: 2001-04-03
; PRIOR APPLICATION NUMBER: US 60/147,133
; PRIOR FILING DATE: 1999-08-04
; NUMBER OF SEQ ID NOS: 23
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 22
; LENGTH: 1141
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: promoter
; LOCATION: (1)..(1141)
; OTHER INFORMATION: consensus sequence of A.t., L.a., and B.n. PAB1 promoters
US-09-806-708B-22

Query Match          3.6%; Score 44; DB 3; Length 1141;
Best Local Similarity 10.7%; Pred. No. 0.02; Mismatches 215; Indels 11; Gaps 2;
Matches 54; Conservative 227;

QY 230 ACCCTGATACGATGAGAAAGAGTATGTTGTTCTCTGACATACCTGACACAA 289
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 53 MSRSKMTWABMYCKYRWRNNNSRWKMGYKKKWBBCANTSBRIHARWKMKTAYEM 112
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 290 TATCTATGTTACTTACACACATTAATGACGCCAACACATTAAGTTATTTGAT 349
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 113 TMTNKGKTCGRHRHYRWRMBDVTVDHYTAAANNNAWTTWCMDKDKRTRMMWKNNNA 172
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 350 AGTATTTGCTGAAGCAGACAGATGCTGCTGAGAAAGTGAATGTTGGTAACCTGA 409
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 173 TGMDDDKTHMMNNNGCBVTMTWVRKYTDWMSBKNNYGMWKNMSYDVITYWVWD 232
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 410 TTTAATAATA---TCAACTGGGTTAAATAAATAAATAATGCACTTAACCTAAAAA 465
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 233 MCRKRYRWRVTRGRRNRNVAVMBTAHRRYNNGWBTBMAAYRRTVNNNNNNKAKCKRA 292
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 466 AACCAATAACCAAGACTTACTTCCCTCGAAAAAGACATTTACAAGGCGTATCT 525
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 293 KYWGNRAAVNSTCTTWKSKTTRKYSWANNCRAGDANKDHKWKMSAAMGYNNNNNN 352
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 526 TAGCCTTTATTTTACAATAAAGAAATAAACAAGTCCCGATATAGCTGTAATTTAT 585
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 353 NMWTYKARHBAWVWHSAMKMHANAHAYSRKWTBVKRTVNNNNNGITMWRKMA 412
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 586 TCCTAAAA-----GAACAGAACTTCACTATGCTTTAAATTAAGTATTAACCTCA 638
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 413 WYWKMDMBGTNNNNNNNGRTYYGWTAKKMMTYTKAKANNCKMRAABDKTCHNNNTW 472
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 639 GATATCTCTGCAAGCTTAGCCTTACAAACGAGCAGACAGACAGAGCCCACTACTCTC 698
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 473 WKAKTYNNNCYWKSMTNKSHRBAALVYTWYMMWBRYAANNNDYMKACTWYKYBV 532
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 699 TAGGAATAATTAAGGCTGCTGCTCTG 725
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 533 CSKMNMYAAMWYKSSMNTSRYYRWK 559
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 8
US-10-104-047-682/c
; Sequence 682, Application US/10104047
; Patent No. 6943241
; GENERAL INFORMATION:
; APPLICANT: HELIX RESEARCH INSTITUTE
; TITLE OF INVENTION: No. 6943241el full length cDNA
; FILE REFERENCE: HI-A0105
; CURRENT APPLICATION NUMBER: US/10/104,047
```

```
US-10-104-047-682
; CURRENT FILING DATE: 2002-03-25
; PRIOR APPLICATION NUMBER:
; PRIOR FILING DATE:
; NUMBER OF SEQ ID NOS: 4096
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 682
; LENGTH: 2308
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-104-047-682

Query Match          3.6%; Score 43.8; DB 3; Length 2308;
Best Local Similarity 47.6%; Pred. No. 0.033; Mismatches 142; Indels 0; Gaps 0;
Matches 129; Conservative 0;

QY 413 AATAATATCAAACTGGGTTAAATAAATAAATAATGTCATTACTTAATAAATAAACCNA 472
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1592 AGTGAAGTAAACTCATCATCAAAAAAAAAAGAGAAGAGAAGAGAGAAACCTTCA 1533
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 473 TACCAACAAGACTTACTTCCCTTGAAAAAGCACAATTACAGAGGCTGATCTTAGCCTT 532
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1532 AATAAACAACTAAGATGATCTTTAAAGATTAGAAAAGCAAGCAAAACCAACCA 1473
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 533 TATATTACAAATTAAGAAATAAACAAGTCCCGATATAGCTGTAATTTATCTTAA 592
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1472 AATTAGTAAAAAGAAATATAAATAATCAGAGTAAATTAATGAATTTGAATGAAGAA 1413
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 593 AGAACGAAACTTCACTATGCTTTAAATAAAGATTAACCTCAGATCTGCAAGC 652
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1412 ACATAGAAAAAGAACATGAATGAACAAAGATGTTGTTTGAAGCAAAACAAACTGAC 1353
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 653 TTAGCTTACAAACGACGACAGACAGACAGA 683
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1352 AACCTTGAGTGAAGTAAAGAAAAAAGAGA 1322
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 9
US-09-949-016-15777/c
; Sequence 15777, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; CURRENT FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 15777
; LENGTH: 80411
; TYPE: DNA
; ORGANISM: Human
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)...(80411)
; OTHER INFORMATION: n = A,T,C or G
US-09-949-016-15777

Query Match          3.5%; Score 42.6; DB 3; Length 80411;
Best Local Similarity 52.5%; Pred. No. 0.53; Mismatches 84; Indels 0; Gaps 0;
Matches 93; Conservative 0;

QY 305 AACAGACAATTATACAGCCAAACCTTAAGTTTATTTATAGTATTTGTTGTTAA 364
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 17141 AATGACATCTTAATCTTATCCATCAAGCTAAAGATGATTTATGATATGTTAT 17082
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
```

QY 365 GGACAGCATGTGAGTCTGAGAAACTGATGTTGTAATTTAATATATCAAA 424  
DB 17081 ACTAAGACATGCCATTTTAAAGAGGGGCTTTCTTCATGAGATTTCAAAATTAAC 17022  
QY 425 CTGGGTTAAATAAAAAATGTCATTAATTAATAAATAAACCATAATACCAACA 481  
DB 17021 TTGGATATCATTAATAATAACTCAATTTATATAGGACCTCAACAAATTAATAA 16965

RESULT 10  
US-09-949-016-12063  
; Sequence 12063, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: C1001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12063  
; LENGTH: 93493  
; TYPE: DNA  
; ORGANISM: Human  
US-09-949-016-12063

Query Match  
Best Local Similarity 3.5%; Score 42.4; DB 3; Length 93493;  
Matches 91; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 454 ACTTAAAAAACCATAATACCAAGCTTTACTTCCCTTGAAAAAGACATTTAC 513  
DB 13347 ACACACACACACCAAGACCAAAACAACTTCTGCTCTAGGACATACATTTT 13406  
QY 514 AAGGCTGATCTTAGCTTTATATTTACATAAAGAAAAATAAACCAAGCTCCGATATAG 573  
DB 13407 AAGAGCTTACTGAGATATGTGACATACATTAACGTACATATTTAAAGTATGATTG 13466  
QY 574 CTGTATTTTATCTTAAAGAAAGAAAGAACTTCACTATGCTTTAAATTA 625  
DB 13467 ACCTGATTAATTTCTTAATTAATGAATGTAACTGAACTTGCTTATTAAGTTA 13518

RESULT 11  
US-09-949-016-13272  
; Sequence 13272, Application US/09949016  
; Patent No. 6812339  
; GENERAL INFORMATION:  
; APPLICANT: VENTER, J. Craig et al.  
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED  
; FILE REFERENCE: C1001307  
; CURRENT APPLICATION NUMBER: US/09/949,016  
; PRIOR FILING DATE: 2000-04-14  
; PRIOR APPLICATION NUMBER: 60/241,755  
; PRIOR FILING DATE: 2000-10-20  
; PRIOR APPLICATION NUMBER: 60/237,768  
; PRIOR FILING DATE: 2000-10-03  
; PRIOR APPLICATION NUMBER: 60/231,498  
; PRIOR FILING DATE: 2000-09-08  
; NUMBER OF SEQ ID NOS: 207012  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 13272  
; LENGTH: 95020  
; TYPE: DNA

; ORGANISM: Human  
US-09-949-016-13272

Query Match  
Best Local Similarity 3.5%; Score 42.4; DB 3; Length 95020;  
Matches 91; Conservative 0; Mismatches 81; Indels 0; Gaps 0;

QY 454 ACTTAAAAAACCATAATACCAAGCTTTACTTCCCTTGAAAAAGACATTTAC 513  
DB 14005 ACACACACACACCAAGACCAAAACAACTTCTGCTCTAGGACATACATTTT 14064  
QY 514 AAGGCTGATCTTAGCTTTATATTTACATAAAGAAATTAACCAAGCTCCGATATAG 573  
DB 14065 AAGAGCTTACTGAGATATGACATACATTAACCTGATATTTAAAGTATGATTG 14124  
QY 574 CTGTATTTTATCTTAAAGAAAGAAAGAACTTCACTATGCTTTAAATTA 625  
DB 14125 ACCTGATTAATTTCTTAATTAATGAATGTAACTGAACTTGCTTATTAAGTTA 14176

RESULT 12  
US-09-662-254B-25/C  
; Sequence 25, Application US/09662254B  
; Patent No. 6933145  
; GENERAL INFORMATION:  
; APPLICANT: Moyer, Richard W.  
; APPLICANT: Li, Yi  
; TITLE OF INVENTION: Materials and Methods for Delivery and Expression of Heterologous  
; FILE REFERENCE: Uf-221C1X1  
; CURRENT APPLICATION NUMBER: US/09/662,254B  
; PRIOR FILING DATE: 2000-09-14  
; PRIOR APPLICATION NUMBER: 09/086,651  
; PRIOR FILING DATE: 1998-05-29  
; PRIOR APPLICATION NUMBER: 60/224,479  
; PRIOR FILING DATE: 2000-08-10  
; NUMBER OF SEQ ID NOS: 80  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 25  
; LENGTH: 50000  
; TYPE: DNA  
; ORGANISM: *Amesbaeta moorei* entomopoxvirus  
US-09-662-254B-25

Query Match  
Best Local Similarity 3.5%; Score 42.2; DB 3; Length 50000;  
Matches 155; Conservative 0; Mismatches 188; Indels 0; Gaps 0;

QY 369 CAGACATGTGAGTCTGAGAAACTGATGTTGTAATTTAATATCAAACTGA 428  
DB 17706 CAGATATATGTTTATGAAAAATATTAATTAATAAATAAATTAATTAATTAAT 17647  
QY 429 GTTAAATTAATAAATAATGTCATTAATTAATAAATAAATAAATAAATAAATAAATAA 488  
DB 17646 TTGAAATTAATAATGATGATGTTGAAGATGAATAAATTAATGAATTAATTAATTA 17587  
QY 489 CTTCCTCTGGAAGACATTTTAAAGGCTGATCTTACCTTTATTTATTAACAATAAG 548  
DB 17586 TTGATTTAAGTTCAACAAATATTAACACATTAATAAATAAATAAATAAATAAATAA 17527  
QY 549 AAAATTAACCAAGCTCCCATATAGCTGTAATTTTATCTTAAAGAACAACTTTCA 608  
DB 17526 ATATTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 17467  
QY 609 CTATGCTTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 668  
DB 17466 ATGTGATTAATTAATCTTATTTTAAATAATTTCAATGAATAAATAAATAAATAAATAA 17407  
QY 669 CAGACAGACACAGAGCCCACTACTCTAGGAAATTAATTA 711  
DB 17406 TTGATATGTAATTAATTTGTTTAAAGAAATATTAATTAATTAATTAATTAATTAAT 17364

```
RESULT 13
US-09-949-016-83375/C
; Sequence 83375, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 83375
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-83375

Query Match          3.4%; Score 42; DB 3; Length 601;
Best Local Similarity 51.0%; Pred. No. 0.053;
Matches 99; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 409 ATTTAATATATCAAACTGGTTAAATATAAATATGCTACTTAAAAAAGC 468
DB 201 ATTCATGAGAAAGAACTTAATGAAAAAAGTTCCATGAAATATATGTA 142
QY 469 CAATATCAACAAGACTTACTTCCCTTGAAAGACATTTACAGGCGTATC 528
DB 141 TAAGTCATATATAAATTTCTAGAAAAGCTTCAGCATACACGAAATTTAT 82
QY 529 CCTTATATTTACATTAAGAAATTAACCAAGTCCCATATAGCTGTAATTT 588
DB 81 TTTAATGATATATTAAGATTAATTAATGCAATTAAGCTTAATTTCT 22
QY 589 TAAAGACAGAAA 602
DB 21 TAAAGAGTTAGAAA 8

RESULT 14
US-09-949-016-83376/C
; Sequence 83376, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 83376
; LENGTH: 601
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-83376

Query Match          3.4%; Score 42; DB 3; Length 601;
```

```
Best Local Similarity 51.0%; Pred. No. 0.053;
Matches 99; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 409 ATTTAATATATCAAACTGGTTAAATATAAATATGCTACTTAAAAAAGC 468
DB 553 ATTCATGAGAAAGAACTTAATGAAAAAAGTTCCATGAAATATATGTA 434
QY 469 CAATATCAACAAGACTTACTTCCCTTGAAAGACATTTACAGGCGTATC 528
DB 493 TAAGTCATATATAAATTTCTAGAAAAGCTTCAGCATACACGAAATTTAT 434
QY 529 CCTTATATTTACATTAAGAAATTAACCAAGTCCCATATAGCTGTAATTT 588
DB 433 TTTAATGATATATTAAGATTAATTAATGCAATTAAGCTTAATTTCT 374
QY 589 TAAAGACAGAAA 602
DB 373 TAAAGAGTTAGAAA 360

RESULT 15
US-09-949-016-14143
; Sequence 14143, Application US/09949016
; Patent No. 6812339
; GENERAL INFORMATION:
; APPLICANT: VENTER, J. Craig et al.
; TITLE OF INVENTION: POLYMORPHISMS IN KNOWN GENES ASSOCIATED
; WITH HUMAN DISEASE, METHODS OF DETECTION AND USES THEREOF
; FILE REFERENCE: CL001307
; CURRENT APPLICATION NUMBER: US/09/949,016
; PRIOR FILING DATE: 2000-04-14
; PRIOR APPLICATION NUMBER: 60/241,755
; PRIOR FILING DATE: 2000-10-20
; PRIOR APPLICATION NUMBER: 60/237,768
; PRIOR FILING DATE: 2000-10-03
; PRIOR APPLICATION NUMBER: 60/231,498
; PRIOR FILING DATE: 2000-09-08
; NUMBER OF SEQ ID NOS: 207012
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 14143
; LENGTH: 143550
; TYPE: DNA
; ORGANISM: Human
US-09-949-016-14143

Query Match          3.4%; Score 42; DB 3; Length 143550;
Best Local Similarity 51.0%; Pred. No. 1.1;
Matches 99; Conservative 0; Mismatches 95; Indels 0; Gaps 0;

QY 409 ATTTAATATATCAAACTGGTTAAATATAAATATGCTACTTAAAAAAGC 468
DB 100795 ATTCATGAGAAAGAACTTAATGAAAAAAGTTCCATGAAATATATGTA 100854
QY 469 CAATATCAACAAGACTTACTTCCCTTGAAAGACATTTACAGGCGTATC 528
DB 100855 TAAGTCATATATAAATTTCTAGAAAAGCTTCAGCATACACGAAATTTAT 100914
QY 529 CCTTATATTTACATTAAGAAATTAACCAAGTCCCATATAGCTGTAATTT 588
DB 100915 TTTAATGATATATTAAGATTAATTAATGCAATTAAGCTTAATTTCT 100974
QY 589 TAAAGACAGAAA 602
DB 100975 TAAAGAGTTAGAAA 100988
```

Search completed: December 12, 2005, 15:23:59  
Job time : 264 secs





Result No.	Score	Query Match	Length	DB	ID	Description
1	1220	100.0	1220	7	US-10-627-075-1	Sequence 1, Appl1
2	276	22.6	2475	9	US-10-764-420-1571	Sequence 1, Appl1
3	238	19.5	240825	3	US-09-790-289-1	Sequence 1, Appl1
4	238	19.5	240825	7	US-10-468-558-1	Sequence 1, Appl1
5	238	19.5	240825	10	US-11-028-982-1	Sequence 1, Appl1
6	213	17.5	2424	7	US-10-317-803-11	Sequence 11, Appl1
7	213	17.5	2424	8	US-10-789-222-34	Sequence 34, Appl1
8	213	17.5	2424	8	US-10-827-759-5	Sequence 5, Appl1
9	211.8	17.4	62705	7	US-10-317-803-4	Sequence 4, Appl1
10	156.6	12.8	3251	5	US-10-044-090-24	Sequence 24, Appl1
11	140.6	11.5	2846	7	US-10-433-793-187	Sequence 187, Appl1
12	139.4	11.4	2846	7	US-10-433-793-186	Sequence 188, Appl1
13	55	4.5	10710	6	US-10-311-455-214	Sequence 866, Appl1
14	54.4	4.5	12658	8	US-10-311-455-214	Sequence 2114, Appl1
15	52.6	4.3	2265	6	US-10-602-494-272	Sequence 272, Appl1
16	51.6	4.2	6823	6	US-10-311-455-117	Sequence 1117, Appl1
17	51.2	4.2	20486	6	US-10-240-485-164	Sequence 164, Appl1
18	51	4.2	3673778	6	US-10-312-84-2	Sequence 2, Appl1
19	50.2	4.1	9964	6	US-10-311-455-72	Sequence 72, Appl1
20	50	4.1	4415	6	US-10-311-455-1879	Sequence 1879, Appl1
21	50	4.1	4415	6	US-10-290-485-155	Sequence 155, Appl1
22	49.8	4.0	15479	7	US-10-257-166-45	Sequence 45, Appl1
23	49.2	4.0	13784	7	US-10-257-166-144	Sequence 144, Appl1

C	24	49	4.0	6127	6	US-10-240-4455-2	Sequence 2, Appl
C	25	49	4.0	8758	6	US-10-311-455-1091	Sequence 1091, Ap
C	26	48.6	4.0	487	3	US-09-918-995-5922	Sequence 5922, Ap
C	27	48.4	4.0	4661	7	US-10-221-613-93	Sequence 93, Appl
C	28	48.2	4.0	2265	8	US-10-602-494-130	Sequence 130, App
C	29	48	3.9	502	3	US-09-814-353-17606	Sequence 17606, A
C	30	48	3.9	5919	6	US-10-311-455-362	Sequence 362, App
C	31	48	3.9	5919	7	US-10-221-613-64	Sequence 64, Appl
C	32	48	3.9	6876	6	US-10-311-455-1138	Sequence 1738, Ap
C	33	48	3.9	8056	8	US-10-477-126-366	Sequence 366, App
C	34	48	3.9	8895	6	US-10-311-455-305	Sequence 305, App
C	35	47.8	3.9	14950	6	US-10-311-455-1229	Sequence 1229, Ap
C	36	47.8	3.9	16724	6	US-10-311-455-1063	Sequence 1063, Ap
C	37	47.8	3.9	16724	6	US-10-240-4455-89	Sequence 89, Appl
C	38	47.6	3.9	1478	5	US-10-198-846-10387	Sequence 10387, A
C	39	47.6	3.9	6103	6	US-10-311-455-1664	Sequence 1664, App
C	40	47.6	3.9	8056	8	US-10-477-126-240	Sequence 240, App
C	41	47.6	3.9	3673778	6	US-10-312-841-1	Sequence 1, Appl
C	42	47.4	3.9	302	3	US-09-814-353-5335	Sequence 5335, Ap
C	43	47.4	3.9	302	3	US-09-814-353-11622	Sequence 11622, A
C	44	47.2	3.9	545	4	US-09-925-065A-113319	Sequence 31319, A
C	45	47.2	3.9	2240	8	US-10-477-126-102	Sequence 102, App

## ALIGNMENTS

```

RESULT 1
US-10-627-075-1
; Sequence 1, Application US/10627075
; Publication No. US2004009193A1
; GENERAL INFORMATION:
; APPLICANT: Livingston et al.
; TITLE OF INVENTION: Composition and Method for Imaging Cells
; FILE REFERENCE: 203667-019
; CURRENT APPLICATION NUMBER: US/10/627,075
; CURRENT FILING DATE: 2003-07-24
; PRIOR APPLICATION NUMBER: 60/398,583
; PRIOR FILING DATE: 2002-07-25
; NUMBER OF SEQ ID NOS: 1
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
; LENGTH: 1220
; TYPE: DNA
; ORGANISM: Mus musculus
US-10-627-075-1

```

Query Match	100.0%;	Score 1220;	DB 7;	Length 1220;
-------------	---------	-------------	-------	--------------

Best Local Similarity 100.0%; Pred. NO. 2.4e-308;  
Matches 1220; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Oy	1	GAGTCTTCCAGATCCGATCTTGCAGCATTTAACTTCACTCATGAAAGGGGTGTCATC	60
Db	1	GAGCTTCCCAATACCCATCTCTGCAGCATTTAACTTCACTCATGAAAGGGGTGTCATC	60
Oy	61	TGGACCAAGAGGCAAGAGCTGGSAAGTCTTTTGAAGTCAGTGCAGCCCCAGCTTTT	120
Db	61	TGGACCAAGAGGCCACAGAGCTGGAAAGTCTTTTGAAGTCAGTGCAGCCCCAGCTTTT	120
Oy	121	ATGCGCAGGGGCTTTTGCMACTTAATTTAAAGGGGAAAGTGATTTTGCCTGAGCCCACTGAC	180
Db	121	ATGCGCAGGGGCTTTTGAACCTTAATTTAAAGGGGAAAGTGATTTTGCCTGAGCCCACTGAC	180
Oy	181	TGGGACCTAATTTTAATCAGGAACATGCCACAGATGATGAGCCCCGAGGAAACCTTGATACA	240
Db	181	TGGGACCTAATTTTAATCAGGAACATGCCACAGATGATGAGCCCCGAGGAAACCTTGATACA	240
Oy	241	GTCGAAGAAAGAGTGATGTTTGTTCCTCTGCACATCTTCACTCAAAATATCTATTGTT	300
Db	241	GTCGAAGAAAGAGTGATGTTTGTTCCTCTGCACATCTTCACTCAAAATATCTATTGTT	300
Oy	301	ACTTAACAGACAAATTAATCAGGCCAAACCACTTTAAAGTTTATTTGTAATGATTTTGG	360

Db 301 ACTTAAGACGACATTAATTAAGCCAAACCACTTAAGTTTATTTGATATGATTTTGTG 360  
QY 361 TTAAGGACAGACATGAGAGCTGAGAGAAAATGATGTTGGTAACTTGATTTATATAT 420  
Db 361 TTAAGGACAGACATGAGAGCTGAGAGAAAATGATGTTGGTAACTTGATTTATATAT 420  
QY 421 CAAACGAGGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 480  
Db 421 CAAACGAGGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 480  
QY 481 AGACTTACTTCCCTTGGAAAAGACATTTACAGGGCTGATCTTATGCTTTATATTTA 540  
Db 481 AGACTTACTTCCCTTGGAAAAGACATTTACAGGGCTGATCTTATGCTTTATATTTA 540  
QY 541 CAATTAAGAAAATTAACCAAGGCTCCGATATAGCTGTAATTTTATTTCTTAAAGAACAG 600  
Db 541 CAATTAAGAAAATTAACCAAGGCTCCGATATAGCTGTAATTTTATTTCTTAAAGAACAG 600  
QY 601 AACTTCTACTATGCTTTAAATTAATTAAGTATTAAGTATTAAGTATTAAGTATTAAG 660  
Db 601 AACTTCTACTATGCTTTAAATTAATTAAGTATTAAGTATTAAGTATTAAGTATTAAG 660  
QY 661 CAAACGAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAG 720  
Db 661 CAAACGAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAGACAG 720  
QY 721 CTCTGACATGCGGAGGAGGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 780  
Db 721 CTCTGACATGCGGAGGAGGCTTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 780  
QY 781 CAGTCTTTGGGGCAGTAAACATTAAGCTTGAATTTTCTGTGTGTGTGTGTGTGTGTGT 840  
Db 781 CAGTCTTTGGGGCAGTAAACATTAAGCTTGAATTTTCTGTGTGTGTGTGTGTGTGTGT 840  
QY 841 CCTACAGGAAGATAGT 900  
Db 841 CCTACAGGAAGATAGT 900  
QY 901 CTCTTATCACTTATATATTAATTAAGGAAAGAAATGATTAATTTGATTAATTTGAT 960  
Db 901 CTCTTATCACTTATATATTAATTAAGGAAAGAAATGATTAATTTGATTAATTTGAT 960  
QY 961 ACTCAGGGGAGAAACAAAGAGTCCGTGACAGACCTCTGAGTGAAGCAGGCTCTCTTCC 1020  
Db 961 ACTCAGGGGAGAAACAAAGAGTCCGTGACAGACCTCTGAGTGAAGCAGGCTCTCTTCC 1020  
QY 1021 TCTCAGGACAGCTCCGAGTGTGCGGGGAGAAAGAGAAAGAGACAGGACCTGGGAAA 1080  
Db 1021 TCTCAGGACAGCTCCGAGTGTGCGGGGAGAAAGAGAAAGAGACAGGACCTGGGAAA 1080  
QY 1081 GAGCCTGCTGCGGGAGCGAGAAAGGCTCTTACATGATGACTTATTAACAGGACAGCCT 1140  
Db 1081 GAGCCTGCTGCGGGAGCGAGAAAGGCTCTTACATGATGACTTATTAACAGGACAGCCT 1140  
QY 1141 GTGCTTTAGACAGCAGCTGAGAGCTCAGAGCCAAAGTTTGAACTCAGATTTAGAAC 1200  
Db 1141 GTGCTTTAGACAGCAGCTGAGAGCTCAGAGCCAAAGTTTGAACTCAGATTTAGAAC 1200  
QY 1201 CCAAAAAGAGAGAGGAATG 1220  
Db 1201 CCAAAAAGAGAGAGGAATG 1220

RESULT 2  
US-10-764-420-1571  
; Sequence 1571, Application US/10764420  
; Publication No. US20050084872A1  
; GENERAL INFORMATION:  
; APPLICANT: Lum, Pek Yee  
; APPLICANT: Tan, Pei Jun  
; APPLICANT: Dai, Hongyue  
; TITLE OF INVENTION: Methods For Determining Whether An Agent Possesses A Defined Biological Activity

FILE REFERENCE: ROSA122057  
; CURRENT APPLICATION NUMBER: US/10/764,420  
; PRIOR FILING DATE: 2004-01-23  
; PRIOR APPLICATION NUMBER: US 60/442,797  
; PRIOR FILING DATE: 2003-01-24  
; PRIOR APPLICATION NUMBER: US 60/474,413  
; PRIOR FILING DATE: 2003-05-30  
; NUMBER OF SEQ ID NOS: 3683  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 1571  
; LENGTH: 2475  
; TYPE: DNA  
; ORGANISM: Mus musculus  
US-10-764-420-1571

Query Match 22.6% Score 276; DB 9; Length 2475;  
Best Local Similarity 100.0%; Pred. No. 4,36-61;  
Matches 276; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 945 GATACTGACACTGTAGACTCAGGAGGAGAAACAAAGAGTCCGTGACAGCTCTGAGAGT 1004  
Db 1 GATACTGACACTGTAGACTCAGGAGGAGAAACAAAGAGTCCGTGACAGCTCTGAGAGT 60  
QY 1005 CAGGCTGCTCTTCTCTCTCAGACAGCTCCGAGTGTCCCGGAGAAAGAGAGAA 1064  
Db 61 CAGGCTGCTCTTCTCTCTCAGACAGCTCCGAGTGTCCCGGAGAAAGAGAGAA 120  
QY 1065 GACAGGACACTGAGAAAGAGCTGCTGCGGAGCGAGAAAGGCTTCACATGATGACT 1124  
Db 121 GACAGGACACTGAGAAAGAGCTGCTGCGGAGCGAGAAAGGCTTCACATGATGACT 180  
QY 1125 CACAGGACAGACCTGCTGCTTGAACAGCAGCTGAGAGCTCAGAGCGCAAGTTGCTGA 1184  
Db 181 CACAGGACAGACCTGCTGCTTGAACAGCAGCTGAGAGCTCAGAGCGCAAGTTGCTGA 240  
QY 1185 ACTCAGCTTTAGAACCCAAAGAGAGAGAGATG 1220  
Db 241 ACTCAGCTTTAGAACCCAAAGAGAGAGAGATG 276

RESULT 3  
US-09-790-289-1/c  
; Sequence 1, Application US/09790289  
; Publication No. US20030165826A1  
; GENERAL INFORMATION:  
; APPLICANT: Caroline Barry  
; APPLICANT: Ilya Chumakov  
; TITLE OF INVENTION: Pg-3 and Biallelic Markers Thereof  
; FILE REFERENCE: 68, US3, REG  
; CURRENT APPLICATION NUMBER: US/09/790,289  
; CURRENT FILING DATE: 2001-02-23  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: Patent .pm  
; SEQ ID NO 1  
; LENGTH: 240825  
; TYPE: DNA  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc feature  
; LOCATION: 1..2000  
; OTHER INFORMATION: 5'regulatory region  
; NAME/KEY: exon  
; LOCATION: 2001..2079  
; OTHER INFORMATION: exon A  
; NAME/KEY: exon  
; LOCATION: 4627..4718  
; OTHER INFORMATION: exon B  
; NAME/KEY: exon  
; LOCATION: 10115..10233  
; OTHER INFORMATION: exon C  
; NAME/KEY: exon  
; LOCATION: 26810..26897  
; OTHER INFORMATION: exon D

NAME/KEY: exon  
LOCATION: 31357..31471  
OTHER INFORMATION: exon E  
NAME/KEY: exon  
LOCATION: 34261..34404  
OTHER INFORMATION: exon F  
NAME/KEY: exon  
LOCATION: 37377..37466  
OTHER INFORMATION: exon S  
NAME/KEY: exon  
LOCATION: 39704..40858  
OTHER INFORMATION: exon T  
NAME/KEY: exon  
LOCATION: 50436..50545  
OTHER INFORMATION: exon G  
NAME/KEY: exon  
LOCATION: 72881..72918  
OTHER INFORMATION: exon H  
NAME/KEY: exon  
LOCATION: 75989..76151  
OTHER INFORMATION: exon I  
NAME/KEY: exon  
LOCATION: 95111..95188  
OTHER INFORMATION: exon J  
NAME/KEY: exon  
LOCATION: 216015..216252  
OTHER INFORMATION: exon K  
NAME/KEY: exon  
LOCATION: 237526..238825  
OTHER INFORMATION: exon L  
NAME/KEY: misc feature  
LOCATION: 238826..240825  
OTHER INFORMATION: 3'regulatory region  
NAME/KEY: allele  
LOCATION: 1999  
OTHER INFORMATION: 5-390-177 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 4601  
OTHER INFORMATION: 5-391-43 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 10228  
OTHER INFORMATION: 5-392-222 : polymorphic base G or T  
NAME/KEY: allele  
LOCATION: 10286  
OTHER INFORMATION: 5-392-280 : polymorphic base G or T  
NAME/KEY: allele  
LOCATION: 10370  
OTHER INFORMATION: 5-392-364 : insertion of G  
NAME/KEY: allele  
LOCATION: 39944  
OTHER INFORMATION: 4-58-318 : polymorphic base G or T  
NAME/KEY: allele  
LOCATION: 39973  
OTHER INFORMATION: 4-58-289 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 41385  
OTHER INFORMATION: 4-54-199 : polymorphic base A or C  
NAME/KEY: allele  
LOCATION: 41404  
OTHER INFORMATION: 4-54-180 : polymorphic base A or C  
NAME/KEY: allele  
LOCATION: 42332  
OTHER INFORMATION: 4-51-312 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 67475  
OTHER INFORMATION: 99-86-266 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 69521  
OTHER INFORMATION: 4-88-107 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 72838  
OTHER INFORMATION: 5-397-141 : polymorphic base G or T  
NAME/KEY: allele

LOCATION: 76060  
OTHER INFORMATION: 5-398-203 : polymorphic base A or C  
NAME/KEY: allele  
LOCATION: 81253  
OTHER INFORMATION: 99-12738-248 : polymorphic base A or C  
NAME/KEY: allele  
LOCATION: 83921  
OTHER INFORMATION: 99-109-358 : polymorphic base A or C  
NAME/KEY: allele  
LOCATION: 91917  
OTHER INFORMATION: 99-12749-175 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 95349  
OTHER INFORMATION: 4-21-154 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 95511  
OTHER INFORMATION: 4-21-317 : polymorphic base G or T  
NAME/KEY: allele  
LOCATION: 96190  
OTHER INFORMATION: 4-23-326 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 97294  
OTHER INFORMATION: 99-12753-34 : polymorphic base A or T  
NAME/KEY: allele  
LOCATION: 98024  
OTHER INFORMATION: 5-364-252 : polymorphic base G or T  
NAME/KEY: allele  
LOCATION: 98914  
OTHER INFORMATION: 99-12755-280 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 98963  
OTHER INFORMATION: 99-12755-329 : polymorphic base A or C  
NAME/KEY: allele  
LOCATION: 103593  
OTHER INFORMATION: 4-87-212 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 104398  
OTHER INFORMATION: 99-12757-318 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 106373  
OTHER INFORMATION: 99-12758-102 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 106407  
OTHER INFORMATION: 99-12758-136 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 108315  
OTHER INFORMATION: 4-105-98 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 108327  
OTHER INFORMATION: 4-105-86 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 108472  
OTHER INFORMATION: 4-45-49 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 109196  
OTHER INFORMATION: 4-44-277 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 114604  
OTHER INFORMATION: 4-86-60 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 115716  
OTHER INFORMATION: 4-84-334 : polymorphic base A or G  
NAME/KEY: allele  
LOCATION: 122083  
OTHER INFORMATION: 99-78-321 : polymorphic base A or T  
NAME/KEY: allele  
LOCATION: 123124  
OTHER INFORMATION: 99-12767-36 : polymorphic base G or C  
NAME/KEY: allele  
LOCATION: 123231  
OTHER INFORMATION: 99-12767-143 : polymorphic base C or T  
NAME/KEY: allele  
LOCATION: 123277

```
OTHER INFORMATION: 99-12767-189 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 123468
OTHER INFORMATION: 99-12767-380 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 126738
OTHER INFORMATION: 4-80-328 : polymorphic base C or T
NAME/KEY: allele
LOCATION: 128210
OTHER INFORMATION: 4-36-384 : polymorphic base G or C
NAME/KEY: allele
LOCATION: 128330
OTHER INFORMATION: 4-36-264 : polymorphic base A or G
NAME/KEY: allele
LOCATION: 128333
OTHER INFORMATION: 4-36-261 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 128594
OTHER INFORMATION: 4-35-333 : polymorphic base A or C
NAME/KEY: allele
LOCATION: 128687
OTHER INFORMATION: 4-35-240 : polymorphic base G or C
NAME/KEY: allele
```

```
Query Match 19.5%; Score 238; DB 3; Length 240825;
Best Local Similarity 64.3%; Pred. No. 4,7e-50;
Matches 545; Conservative 0; Mismatches 265; Indels 38; Gaps 11;
```

```
QY 221 CCCGAGAAACCTGATACGTAGAGAAAGGTGATGTTGTTCTCTCGACATAGT 280
DB 158572 CTCMAAAAAAAAAAAAAAAAAAAAAAAAAAGTTGTAATCCTCTTCTACAGACACTT 158513
QY 281 TCACGCAATATCTATGTTACTTACAGACATTAATCGGCCAACCACTTAAGTTT 340
DB 158512 AAATTTCTAATTTCTCTATGTAATGTAAGACTTAATCTAGTATGTTTGAAGT 158453
QY 341 TATTTGTATAGTATTTTGTGTAAGCACAGACATGATGCTGAGAAAATGATGTTG 400
DB 158452 GAATATCCAAAGTGAGTTTTTTTAAAGA-ATAAAGTCAGAGCTCAG-AACTGATGTTT 158395
QY 401 GTAACCTGATTTAATA-----TATCAACTGGGTTAAATMAAA-----AA 443
DB 158394 CTAAAGCTGATTAATAACAGATCTTTCTATCAACTGTTCCGAAATGAAATCTTTTGT 158335
QY 444 AATGTCATTAACCTTAATAAAAAAAAAACCAATACCAACAGACTTTACTCCCTGGA 503
DB 158334 ACTTGCTAGTGTGTTCAAGATCACTAAATCAATGAATCTATTTTCTCTGAAAG 158275
QY 504 GCACATTTCAAGGG--CTGATCTTAGCCTTTATATTATTAATAAAGAAATAACCA 560
DB 158274 GCACACGTTTACAGAGCCCAACTCTCTCTTATTTGTAATAACAAATAAACGA 158215
QY 561 GGTCCGATATAGCTGTATTTTATTCCTA-----AAAGAACGAAACTTTCATG 613
DB 158214 AGTCTGACCTATTTGTATATTTTATTTCTTAAGGAAACAGAACTTTCATTTGA 158155
QY 614 CTTTAAATTAAGTATTAACCTCAAGTACTCTGCAAGCTTGAAGCTTACAAAGAGAG-- 671
DB 158154 CTTCAACATTAAGTTATTAACCTCAGATATTTTGCAGCTTGACACGGCAAAATCACT 158095
QY 672 ACAGCAACAGAGCCCA--GCTACTCTCTAGGAAA---TAATTAGGGTGTGCTCTGA 726
DB 158094 TCAGCAAAAGAGATCACTGCTCTCTTAGGAAATACTTAATTGGGGTGTGCTTAGGA 158035
QY 727 CATGCCAGGGGCTTGTGCTGTGCTG- TGTTCCAGAGAGCTCTGACAGTACACACTC 785
DB 158034 AATGCCAGGGGCTCTGTAAAGATCGGTTTTTCCAGAGGGTCTCTCAGAGATGGGCTC 157975
QY 786 CTTTGGGCACTGAACACTATGCTGTGATTTTCTGTTGCTGCTGCTAGTACCCCTAC 845
DB 157974 TGGTTGAGGGGAGGCACTTGTCTGTGATTTTCTGTTGCTGCTGCTAGTACCCCTAC 157915
QY 846 AGAAGATAGTGGTGAAGCAAGG-GGGGGAGCGGCTGGCTGACATGTCTGGCTGCTCT 904
```

```
DB 157914 AGAAGATATACGGCTTAAGCCAGAGGGCCGAGACGCCCACTACATGTGTGCTGCT 157855
QY 905 TATCAACTTATCATATTAAGGAAAGAAAGTGAATTTGATTTGATTAAGTACTGACACTGTA 964
DB 157854 TATCAACTTATCATATTAAGGAAAGAAAGTGAATTTGATTTGATTAAGTACTGACACTGTA 157795
QY 965 AGGAGAAACAAAGAGTCCGTGACAGACTCTGAGTGAAGAGGGCTGCTCTTCTCTC 1024
DB 157794 TGGGAGAGAGAAACAAAGAGCCGTGAAGCTGCTCTGTAAGAGCTGACACAGCCCTCC 157735
QY 1025 AGGACAGC 1032
DB 157734 AAGTGAGC 157727

RESULT 4
US-10-468-582-1/c
Sequence 1, Application US/10468582
Publication No. US20040163137A1
GENERAL INFORMATION:
APPLICANT: Genet
TITLE OF INVENTION: PG-3 and diallelic markers thereof
FILE REFERENCE: 68,WO2
CURRENT APPLICATION NUMBER: US/10/468,582
CURRENT FILING DATE: 2003-08-20
NUMBER OF SEQ ID NOS: 5
SOFTWARE: Patent.pm
SEQ ID NO 1
LENGTH: 240825
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc feature
LOCATION: 1..2000
OTHER INFORMATION: 5' regulatory region
FEATURE:
NAME/KEY: exon
LOCATION: 2001..2079
OTHER INFORMATION: exon A
FEATURE:
NAME/KEY: exon
LOCATION: 4627..4718
OTHER INFORMATION: exon B
FEATURE:
NAME/KEY: exon
LOCATION: 10115..10233
OTHER INFORMATION: exon C
FEATURE:
NAME/KEY: exon
LOCATION: 26810..26897
OTHER INFORMATION: exon D
FEATURE:
NAME/KEY: exon
LOCATION: 31357..31471
OTHER INFORMATION: exon E
FEATURE:
NAME/KEY: exon
LOCATION: 34261..34404
OTHER INFORMATION: exon F
FEATURE:
NAME/KEY: exon
LOCATION: 37377..37466
OTHER INFORMATION: exon S
FEATURE:
NAME/KEY: exon
LOCATION: 39704..40858
OTHER INFORMATION: exon T
FEATURE:
NAME/KEY: exon
LOCATION: 50436..50545
OTHER INFORMATION: exon G
FEATURE:
```

NAME/KEY: exon  
LOCATION: 72881..72918  
OTHER INFORMATION: exon H  
FEATURE:  
NAME/KEY: exon  
LOCATION: 75989..76151  
OTHER INFORMATION: exon I  
FEATURE:  
NAME/KEY: exon  
LOCATION: 95111..95188  
OTHER INFORMATION: exon J  
FEATURE:  
NAME/KEY: exon  
LOCATION: 216015..216252  
OTHER INFORMATION: exon K  
FEATURE:  
NAME/KEY: exon  
LOCATION: 237526..238825  
OTHER INFORMATION: exon L  
FEATURE:  
NAME/KEY: misc\_feature  
LOCATION: 238826..240825  
OTHER INFORMATION: 3'regulatory region  
FEATURE:  
NAME/KEY: allele  
LOCATION: 1999  
OTHER INFORMATION: 5-390-177 : polymorphic base G or C  
FEATURE:  
NAME/KEY: allele  
LOCATION: 4601  
OTHER INFORMATION: 5-391-43 : polymorphic base A or G  
FEATURE:  
NAME/KEY: allele  
LOCATION: 10228  
OTHER INFORMATION: 5-392-222 : polymorphic base G or T  
FEATURE:  
NAME/KEY: allele  
LOCATION: 10286  
OTHER INFORMATION: 5-392-280 : polymorphic base G or T  
FEATURE:  
NAME/KEY: allele  
LOCATION: 10370  
OTHER INFORMATION: 5-392-364 : insertion of G  
FEATURE:  
NAME/KEY: allele  
LOCATION: 39973  
OTHER INFORMATION: 4-58-289 : polymorphic base G or C  
FEATURE:  
NAME/KEY: allele  
LOCATION: 41385  
OTHER INFORMATION: 4-54-199 : polymorphic base A or C  
FEATURE:  
NAME/KEY: allele  
LOCATION: 41404  
OTHER INFORMATION: 4-54-180 : polymorphic base A or C  
FEATURE:  
NAME/KEY: allele  
LOCATION: 42232  
OTHER INFORMATION: 4-51-312 : polymorphic base G or C  
FEATURE:  
NAME/KEY: allele  
LOCATION: 67475  
OTHER INFORMATION: 99-86-266 : polymorphic base A or G  
FEATURE:  
NAME/KEY: allele  
LOCATION: 69521  
OTHER INFORMATION: 4-88-107 : polymorphic base A or G  
FEATURE:  
NAME/KEY: allele

LOCATION: 72838  
OTHER INFORMATION: 5-397-141 : polymorphic base G or T  
FEATURE:  
NAME/KEY: allele  
LOCATION: 76060  
OTHER INFORMATION: 5-398-203 : polymorphic base A or C  
FEATURE:  
NAME/KEY: allele  
LOCATION: 81253  
OTHER INFORMATION: 99-12738-248 : polymorphic base A or C  
FEATURE:  
NAME/KEY: allele  
LOCATION: 83921  
OTHER INFORMATION: 99-109-358 : polymorphic base A or C  
FEATURE:  
NAME/KEY: allele  
LOCATION: 91917  
OTHER INFORMATION: 99-12749-175 : polymorphic base C or T  
FEATURE:  
NAME/KEY: allele  
LOCATION: 95349  
OTHER INFORMATION: 4-21-154 : polymorphic base C or T  
FEATURE:  
NAME/KEY: allele  
LOCATION: 95511  
OTHER INFORMATION: 4-21-317 : polymorphic base G or T  
FEATURE:  
NAME/KEY: allele  
LOCATION: 96190  
OTHER INFORMATION: 4-23-326 : polymorphic base A or G  
FEATURE:  
NAME/KEY: allele  
LOCATION: 97294  
OTHER INFORMATION: 99-12753-34 : polymorphic base A or T  
FEATURE:  
NAME/KEY: allele  
LOCATION: 98024  
OTHER INFORMATION: 5-364-252 : polymorphic base G or T  
FEATURE:  
NAME/KEY: allele  
LOCATION: 98914  
OTHER INFORMATION: 99-12755-280 : polymorphic base A or G  
FEATURE:  
NAME/KEY: allele  
LOCATION: 98963  
OTHER INFORMATION: 99-12755-329 : polymorphic base A or C  
FEATURE:  
NAME/KEY: allele  
LOCATION: 103593  
OTHER INFORMATION: 4-87-212 : polymorphic base A or G  
FEATURE:  
NAME/KEY: allele  
LOCATION: 104398  
OTHER INFORMATION: 99-12757-318 : polymorphic base C or T  
FEATURE:  
NAME/KEY: allele  
LOCATION: 106373  
OTHER INFORMATION: 99-12758-102 : polymorphic base A or G  
FEATURE:  
NAME/KEY: allele  
LOCATION: 106407  
OTHER INFORMATION: 99-12758-136 : polymorphic base C or T  
FEATURE:  
NAME/KEY: allele  
LOCATION: 108315  
OTHER INFORMATION: 4-105-98 : polymorphic base A or G  
FEATURE:  
NAME/KEY: allele  
LOCATION: 108327  
OTHER INFORMATION: 4-105-86 : polymorphic base A or G  
FEATURE:  
NAME/KEY: allele

Query Match 19.5%; Score 238; DB 7; Length 240825;  
Best Local Similarity 64.3%; Pred. No. 4.7e-50;  
Matches 545; Conservative 0; Mismatches 265; Indels 38; Gaps 11;

QY 221 CCCGAGGAAACCCGTGATACAGTGAAGGAAAGGTAATGTTGTTCCCTCGACATCT 280  
DB 158572 CTCAGAAAAAAGGTAATGTTGTTCCCTCGACATCT 280  
QY 281 TCACCAATATCTATCTTGTCTTAAACAGCAATTAATCAGGCCAAACCTTTAACTTT 340  
DB 158512 AAATCTATTTCTCTATTTGTAATGTAATGAACCTTAATCAGTACAGTATTTTAAAGT 158453  
QY 341 TATTTGTAATGTAATTTTGTGTAAGGCAAGACATGTAGTGTGAGAAACGATGTG 400  
DB 158452 GAATATCCAGTGAAGTATTTTAAAGA-ATTAAGTCAAGCTCAG-AACTGATGTTT 158395  
QY 401 GTAACTTGATTTAATA-----TATCAAACTGGGTTAAATAAAAA-----AA 443  
DB 158394 CTAAAGTACTTAATAACAGATCTTTCTATCCAACTGTTGGAATGAAAACTTTTGT 158335  
QY 444 AATGTCATACTTAATAAAAAAACCATAACCAAGAACTTTACTTCCCTTGAAAA 503  
DB 158334 ACTTGGCTAGTGTCTTCAAGAAATCACTAAATCAATGATCTATTTTCCCTGAAAA 158275  
QY 504 GCACATTTACAAGG--CTGATCTTAACCTTTATTTTACATAAAGAAATTAACCA 560  
DB 158274 GCACACGTTTACAGAGAGCCAAACTTCTCTTATTTTGAATAAATAAATTAACCA 158215  
QY 561 GGTCGCCGATATAGCTGTAATTTTATTCCTA-----AAAGAACAAACTTTCACTAG 613  
DB 158214 AGTCTGAACCTATTTGTAATTTTATTTTCTTAAAGAAAAACAGGACCTTTCACTGTA 158155  
QY 614 CTTTAAATTAAGTATTAACCTGATTAACCTGATTAACCTTAAGCTTAACCAAGAGAG-- 671  
DB 158154 CTTCAACATTAAGTATTAACCTGATTAACCTGATTAACCTTAAGCTTAACCAAGAGAG-- 158095  
QY 672 ACAGACAAGAGGCCCA-GCTACTCTCTGAAAA---TAATTAAGGTGGTCCCTGTA 726  
DB 158094 TCAGACAAAGAGATCAAGCTCTCTCTAGAAAACTTAATTAAGGTGGTCCCTGTA 158035  
QY 727 CATGCCAGGAGGCTTGTGCTGCTGCTG--TGTTCCCAAGAGGCTTCTGCAATGACAGTC 785  
DB 158034 AATGCCAGGAGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 157975  
QY 786 CTTTGGGCGATGAAGCACTATGCTGATTTTCTGTTGCTGCTGCTGCTGCTGCTGCT 845  
DB 157974 TGGTTGAAGGCGAGGCAATCTGCTGATTTTCTGTTGCTGCTGCTGCTGCTGCT 157915  
QY 846 AGGAAGATAGTGGTGGAGGAGG--GGGCGAGGCGGCTGCTGCTGCTGCTGCTGCT 904  
DB 157914 AGGAAGATAGTGGTGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 157855  
QY 905 TATCAACTTATCTATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 964  
DB 157854 TATCAACTTATCTATTAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 157795  
QY 965 AGGGAGGAGAAACAAGAGTCCGTCAGACCTCTGAGTGAAGCAGGAGGCTGCTTCTCTC 1024  
DB 157794 TGGGGAGAGAGGAAACAAGAGCCTGTAAGGCTGCTGTAAGAGCTGACACAGCCTCC 157735  
QY 1025 AGGACAGC 1032  
DB 157734 AAGTGAGC 157727

RESULT 5  
US-11-028-971-1/c  
; Sequence 1, Application US/11028971  
; Publication No. US20050158779A1  
; GENERAL INFORMATION:  
; APPLICANT: Caroline Barry  
; APPLICANT: Ilya Chumakov  
; TITLE OF INVENTION: PG-3 and Biallelic Markers Thereof

FILE REFERENCE: 68, US3, REG  
CURRENT APPLICATION NUMBER: US/11/028, 971  
CURRENT FILING DATE: 2005-01-04  
PRIOR APPLICATION NUMBER: US/09/790, 289  
PRIOR FILING DATE: 2001-02-23  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: Patent.pm  
SEQ ID NO 1  
LENGTH: 240825  
TYPE: DNA  
ORGANISM: Homo sapiens  
FEATURE:  
NAME/KEY: misc.feature  
LOCATION: 1..2000  
OTHER INFORMATION: 5' regulatory region  
FEATURE:  
NAME/KEY: exon  
LOCATION: 2001..2079  
OTHER INFORMATION: exon A  
FEATURE:  
NAME/KEY: exon  
LOCATION: 4627..4718  
OTHER INFORMATION: exon B  
FEATURE:  
NAME/KEY: exon  
LOCATION: 10115..10233  
OTHER INFORMATION: exon C  
FEATURE:  
NAME/KEY: exon  
LOCATION: 26810..26897  
OTHER INFORMATION: exon D  
FEATURE:  
NAME/KEY: exon  
LOCATION: 31357..31471  
OTHER INFORMATION: exon E  
FEATURE:  
NAME/KEY: exon  
LOCATION: 34261..34404  
OTHER INFORMATION: exon F  
FEATURE:  
NAME/KEY: exon  
LOCATION: 37377..37466  
OTHER INFORMATION: exon S  
FEATURE:  
NAME/KEY: exon  
LOCATION: 39704..40858  
OTHER INFORMATION: exon T  
FEATURE:  
NAME/KEY: exon  
LOCATION: 50436..50545  
OTHER INFORMATION: exon G  
FEATURE:  
NAME/KEY: exon  
LOCATION: 72881..72918  
OTHER INFORMATION: exon H  
FEATURE:  
NAME/KEY: exon  
LOCATION: 75989..76151  
OTHER INFORMATION: exon I  
FEATURE:  
NAME/KEY: exon  
LOCATION: 95111..95188  
OTHER INFORMATION: exon J  
FEATURE:  
NAME/KEY: exon  
LOCATION: 216015..216252  
OTHER INFORMATION: exon K  
FEATURE:  
NAME/KEY: exon  
LOCATION: 237526..238825  
OTHER INFORMATION: exon L  
FEATURE:  
NAME/KEY: misc.feature

```

LOCATION: 238826..240825
OTHER INFORMATION: 3/regulatory region
FEATURE:
NAME/KEY: allele
LOCATION: 1999
OTHER INFORMATION: 5-390-177 : polymorphic base G or C
FEATURE:
NAME/KEY: allele
LOCATION: 4601
OTHER INFORMATION: 5-391-43 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 10228
OTHER INFORMATION: 5-392-222 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 10286
OTHER INFORMATION: 5-392-280 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 10370
OTHER INFORMATION: 5-392-364 : insertion of G
FEATURE:
NAME/KEY: allele
LOCATION: 39944
OTHER INFORMATION: 4-58-318 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 39973
OTHER INFORMATION: 4-58-289 : polymorphic base G or C
FEATURE:
NAME/KEY: allele
LOCATION: 41385
OTHER INFORMATION: 4-54-199 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 41404
OTHER INFORMATION: 4-54-180 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 42232
OTHER INFORMATION: 4-51-312 : polymorphic base G or C
FEATURE:
NAME/KEY: allele
LOCATION: 67475
OTHER INFORMATION: 99-86-266 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 69521
OTHER INFORMATION: 4-88-107 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 72838
OTHER INFORMATION: 5-397-141 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 76060
OTHER INFORMATION: 5-398-203 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 81253
OTHER INFORMATION: 99-12738-248 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 83921
OTHER INFORMATION: 99-109-358 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 91917
OTHER INFORMATION: 99-12749-175 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 95349

```

```

OTHER INFORMATION: 4-21-154 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 95511
OTHER INFORMATION: 4-21-317 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 96190
OTHER INFORMATION: 4-23-326 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 97294
OTHER INFORMATION: 99-12753-34 : polymorphic base A or T
FEATURE:
NAME/KEY: allele
LOCATION: 98024
OTHER INFORMATION: 5-364-252 : polymorphic base G or T
FEATURE:
NAME/KEY: allele
LOCATION: 98914
OTHER INFORMATION: 99-12755-280 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 98963
OTHER INFORMATION: 99-12755-329 : polymorphic base A or C
FEATURE:
NAME/KEY: allele
LOCATION: 103593
OTHER INFORMATION: 4-87-212 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 104398
OTHER INFORMATION: 99-12757-318 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 106373
OTHER INFORMATION: 99-12758-102 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 106407
OTHER INFORMATION: 99-12758-136 : polymorphic base C or T
FEATURE:
NAME/KEY: allele
LOCATION: 108315
OTHER INFORMATION: 4-105-98 : polymorphic base A or G
FEATURE:
NAME/KEY: allele
LOCATION: 108327

```

Query Match 19.5%; Score 238; DB 10; Length 240825;  
 Best Local Similarity 64.3%; Pred. No. 4.7e-50;  
 Matches 545; Conservative 0; Mismatches 265; Indels 38; Gaps 11;

```

QY 221 CCCGAGAAACCTGATACAGTGAAGAAAGGTGATGTTGTTCTCTCGACATACT 280
DB 158572 CTCAAAAAAGTGTGATTCCTCTTCTACAGACACTT 158513
QY 281 TCACCTCAATATCTATTGTTACTTAAACAGACATTAATCAGGCAACACTTAACTT 340
DB 158512 AAATTCATATTTCTCTATTGTATGTAATGAACCTTAATAGACAGTATTTTGAAGT 158453
QY 341 TATTTGTATGATTTTGTGTTAAGCACAGACATGTGAGTCTGAGAAACTGATGTTG 400
DB 158452 GAATATCCAAAGTGAGTTTAAAAAGA-ATAAAGGTGAGACTCG-AAACTGATGTTT 158395
QY 401 GTAACCTGATTTAATAA-----TATCAACTGGGTTAAATAA-----AA 443
DB 158394 CTAAAGTGACTTAATAACAGATCTTCTATCCAGCTTTCGAAATGAAAATCTTTGTA 158335
QY 444 AATGTCATTAACCTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 503
DB 158334 ACTTGCTAGTGCTTCAAGAAATCAACATAAATCAATGATCTATTTTCCCTGAAAAG 158275

```

```
Qy 504 GCACATTACAGG---CTGATCTAGACCTTTATTTACATTAAGAAATTAACCA 560
Db 158274 GCACACGTTTACAGAGCAAACTTCTTCTTTATTTGTAATAACAAATTAACCGA 158215
Qy 561 GGTCCCGATATAGCTGTATTTTATTCCTA-----AAGAACAGAACTTTCACTATG 613
Db 158214 AGTCCACACCAATTTGATATTTTATTTCTTAAGAAAAAACAGAACCTTTCACTATG 158155
Qy 614 CTTTAAATTAATGATTTACCTCAGATACCTCTGCAAGCTTACGCTTACAAACGACG-- 671
Db 158154 CTTCAACCTTAACTTATTAATTAATTTTCCAGATATTTTCCAGCTTACGACGCAAAATCAGTT 158095
Qy 672 ACAGACAAACAGACCCCA-GCTACTCTAGAGAAA-----TAATTAGGGTGTGCTCTGA 726
Db 158094 TCAGACAAAGAGATCAACGCTCTCTCTAGAGAAATCTTAATTGGGGTGTGCTCTAGGA 158035
Qy 727 CATGCCAAGGGCTTTTGTGCTGTG--TGTTCCAGAAAGGCTTCTGACATACAGATC 785
Db 158034 AATGCCAGGGGCTCTCTTAACATCGGTTTTTCCAGAGGTTTCTGACATAGGCTC 157975
Qy 786 CTTTGGGAGTAAAGCACTATGCTGATTTTCCGTTGCTGCTGCTAGTACCCCTTAC 845
Db 157974 TGCTTGAAGGCGACGCACTTCTGCTGATTTTCTGCTGCTGCTAGTACCCCTTAC 157915
Qy 846 AGAAGATATAGTGGTGAAGCCAGG-GGGCGAGCGGCTGCTGACATGTCTGCTGCTCT 904
Db 157914 AGGAAATATACGGCTTAAGCCAGAGGCGGAGGACCACTACATGCTGCTGCTCT 157855
Qy 905 TATCACTTATCTATTAAGAGAAAGAAAGTATTTGAGATCTGACATCTGTAGACTC 964
Db 157854 TATCACTTATCTATTAAGAGAAAGAAAGTATTTGAGATCTGACATCTGTAGACTC 157795
Qy 965 AGGGAGAAACAAAGAGCTCGTGCAAGCTCTGAGTGAAGCAGGCGTCTCTTCTCTC 1024
Db 157794 TGGGAGAGAGAAACAAAGAGCCGTGAAGCTCTCTGTAAAAGCTGACACAGCCCTCC 157735
Qy 1025 AGGACAGC 1032
Db 157734 AAGTGAGC 157727

RESULT 6
US-10-317-803-11
; Sequence 11, Application US/10317803
; Publication No. US20040115640A1
; GENERAL INFORMATION:
; APPLICANT: Kathleen Myers
; APPLICANT: Kenneth W. Doble
; TITLE OF INVENTION: MODULATION OF ANGIOPOIETIN-2 EXPRESSION
; FILE REFERENCE: RTS-0454
; CURRENT APPLICATION NUMBER: US/10/317,803
; CURRENT FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 244
; SEQ ID NO 11
; LENGTH: 2424
; TYPE: DNA
; ORGANISM: M. musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (211)...(1701)
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2308
; OTHER INFORMATION: n = A,T,C or G
US-10-317-803-11

Query Match 17.5%; Score 213; DB 7; Length 2424;
Best Local Similarity 100.0%; Pred. No. 1.3e-44;
Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1008 GCGTCTCTCTCTCTCTAGAGACAGCTCCGAGTGTGCCGGGAGAGAGAGAGAGAC 1067
Db 1 GCGTCTCTCTCTCTCTAGAGACAGCTCCGAGTGTGCCGGGAGAGAGAGAGAGAGAC 60
```

```
Qy 1068 AGGCACGTGGGAAAGAGCGCTGCGGGAGCGAGAAAGCTCTCACTGATGACTTATTC 1127
Db 61 AGGCACGTGGGAAAGAGCGCTGCGGGAGCGAGAAAGCTCTCACTGATGACTTATTC 1120
Qy 1128 AGGCACAGCCCTGTGCTTGAACAGACAGCTGAGAGCTCAGACCAAGTTGCTGA 1187
Db 121 AGGCACAGCCCTGTGCTTGAACAGAGCTGAGAGCTCAGACCAAGTTGCTGA 180
Qy 1188 CACAGTTTGAACCCCAAAAGAGAGAGAAATG 1220
Db 181 CACAGTTTGAACCCCAAAAGAGAGAGAAATG 213

RESULT 7
US-10-789-222-34
; Sequence 34, Application US/10789222
; Publication No. US20040186054A1
; GENERAL INFORMATION:
; APPLICANT: Yu, Qin
; TITLE OF INVENTION: Angiopietin and Fragments, Mutants, and Analogs Thereof and Uses
; FILE REFERENCE: of the Same
; FILE REFERENCE: UPN0003-100 (P3115)
; CURRENT APPLICATION NUMBER: US/10/789,222
; PRIOR FILING DATE: 2004-02-27
; PRIOR APPLICATION NUMBER: US 60/450,582
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 34
; LENGTH: 2424
; TYPE: DNA
; ORGANISM: mouse
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (2308)..(2308)
; OTHER INFORMATION: n is a, c, g, or t
US-10-789-222-34

Query Match 17.5%; Score 213; DB 8; Length 2424;
Best Local Similarity 100.0%; Pred. No. 1.3e-44;
Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1008 GCGTCTCTCTCTCTCTAGAGACAGCTCCGAGTGTGCCGGGAGAGAGAGAGAGAC 1067
Db 1 GCGTCTCTCTCTCTCTAGAGACAGCTCCGAGTGTGCCGGGAGAGAGAGAGAGAGAC 60
Qy 1068 AGGCACGTGGGAAAGAGCGCTGCGGGAGCGAGAAAGCTCTCACTGATGACTTATTC 1127
Db 61 AGGCACGTGGGAAAGAGCGCTGCGGGAGCGAGAAAGCTCTCACTGATGACTTATTC 120
Qy 1128 AGGCACAGCCCTGTGCTTGAACAGACAGCTGAGAGCTCAGACCAAGTTGCTGA 1187
Db 121 AGGCACAGCCCTGTGCTTGAACAGAGCTGAGAGCTCAGACCAAGTTGCTGA 180
Qy 1188 CACAGTTTGAACCCCAAAAGAGAGAGAAATG 1220
Db 181 CACAGTTTGAACCCCAAAAGAGAGAGAAATG 213

RESULT 8
US-10-827-759A-5
; Sequence 5, Application US/10827759A
; Publication No. US20040248174A1
; GENERAL INFORMATION:
; APPLICANT: The Trustees of the University of Pennsylvania
; APPLICANT: Samuel Jotham Reich
; APPLICANT: Michael J. Tolentino
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR siRNA
; FILE REFERENCE: 43826-0005 US1
; CURRENT APPLICATION NUMBER: US/10/827,759A
; CURRENT FILING DATE: 2004-04-19
```



```
; PRIOR APPLICATION NUMBER: US 60/463,981
; PRIOR FILING DATE: 2003-04-18
; NUMBER OF SEQ ID NOS: 736
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 2424
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2308
; OTHER INFORMATION: n = a, g, c or t
US-10-827-759A-5

Query Match      17.5%; Score 213; DB 8; Length 2424;
Best Local Similarity 100.0%; Pred. No. 1.3e-44;
Matches 213; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1008 GGCTGCTCTTCTCTCAGACAGCTCCAGGTGTGCGGGAGAAAGAAAGAGAGAC 1067
DB 1 GGCTGCTCTTCTCTCAGACAGCTCCAGGTGTGCGGGAGAAAGAAAGAGAGAC 60
QY 1068 AGGCACTGGAAAGAGGCTCTGCGGAGCGAGAGAGGCTCTCACTGATGACTTATTTCAC 1127
DB 61 AGGCACTGGAAAGAGGCTCTGCGGAGCGAGAGAGGCTCTCACTGATGACTTATTTCAC 120
QY 1128 ACGGCAAGGCTGTGCTTGAAGACAGAGCTGAGAGCTCAGAGCGAAGTTTGCTGAACT 1187
DB 121 ACGGCAAGGCTGTGCTTGAAGACAGAGCTGAGAGCTCAGAGCGAAGTTTGCTGAACT 180
QY 1188 CACAGTTTAGAACCCCAAAAAGAGAGAGAAATG 1220
DB 181 CACAGTTTAGAACCCCAAAAAGAGAGAGAAATG 213

RESULT 9
US-10-317-803-4
; Sequence 4, Application US/10317803
; Publication No. US20040115640A1
; GENERAL INFORMATION:
; APPLICANT: Kathleen Myers
; APPLICANT: Kenneth W. Doble
; TITLE OF INVENTION: MODULATION OF ANGIOPOIETIN-2 EXPRESSION
; FILE REFERENCE: RTS-0454
; CURRENT APPLICATION NUMBER: US/10/317,803
; CURRENT FILING DATE: 2002-12-11
; NUMBER OF SEQ ID NOS: 244
; SEQ ID NO 4
; LENGTH: 62705
; TYPE: DNA
; ORGANISM: H. sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2308
; OTHER INFORMATION: n = a, g, c or t
US-10-317-803-4

Query Match      17.4%; Score 211.8; DB 7; Length 62705;
Best Local Similarity 74.3%; Pred. No. 1.6e-43;
Matches 335; Conservative 0; Mismatches 107; Indels 9; Gaps 5;

QY 591 AAAGAAGAGAACTTCACTATGCTTAAATTAAGTATTAACCTCAGATCTGCA 650
DB 31 AAAGAAGAGAACTTCACTATGCTTAAATTAAGTATTAACCTCAGATCTGCA 90
QY 651 GCTTACCTTCAAAACGAGCAG--ACAGACAACAGAGCCCCA--GCTACTCTTGAAGAA-- 705
DB 91 GCTTACCTTCAAAACGAGCAG--ACAGACAACAGAGCCCCA--GCTACTCTTGAAGAA-- 150
QY 706 --TAAATTAAGTGTGCTCTTGAATGCTTGAAGGCTTGTGCTGTGCTG--TGTTCACA 762
DB 151 CTTAAATTAAGTGTGCTCTTGAATGCTTGAAGGCTTGTGCTGTGCTG--TGTTCACA 210
QY 763 GAAGGCTTGTGAGTACAGAGTCTTGTGGGGCAGTAAAGCACTATGCTGATTTTCCG 822
DB 211 GAAGGCTTGTGAGTACAGAGTCTTGTGGGGCAGTAAAGCACTATGCTGATTTTCCG 270
```

```
QY 823 TTGCTGTAGTGAAGCCCTTACAGAGAAATAGTGGTAGACGAG--GGCGGAGCGCT 881
DB 271 TTGCTGTAGTGAAGCCCTTACAGAGAAATAGTGGTAGACGAG--GGCGGAGCGCT 330
QY 882 GGCTGACATGTGTGCTGTCTTATCACTTATCATTAAGGAAAGAAAGTATGAT 941
DB 331 CACTACACATGTGTGCTGTCTTATCACTTATCATTAAGGAAAGAAAGTATGAT 390
QY 942 TCGATATCTGACACTGTAGACTCAGGGAGAAACAAAGAGTCCGTGACAGCTCTGAGT 1001
DB 391 TCGATATCTGACACTGTAGACTCAGGGAGAAACAAAGAGTCCGTGAGTCTCT 450
QY 1002 GAGCAGGCTGTCTCTTCTCTCAGAGAC 1032
DB 451 GTRAAAGCTGACACAGCCCTCCCAAGTAGAC 481

RESULT 10
US-10-044-090-24
; Sequence 24, Application US/10044090
; Publication No. US20020137081A1
; GENERAL INFORMATION:
; APPLICANT: Olga Bandman
; TITLE OF INVENTION: GENES DIFFERENTIALLY EXPRESSED IN VASCULAR TISSUE ACTIVATION
; FILE REFERENCE: PA-0028 US
; CURRENT APPLICATION NUMBER: US/10/044,090
; CURRENT FILING DATE: 2002-01-09
; NUMBER OF SEQ ID NOS: 850
; SOFTWARE: PERL Program
; SEQ ID NO 24
; LENGTH: 3251
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 2308
; OTHER INFORMATION: Incyte ID No. US20020137081A1 343965.2
US-10-044-090-24

Query Match      12.8%; Score 156.6; DB 5; Length 3251;
Best Local Similarity 74.9%; Pred. No. 8.9e-30;
Matches 209; Conservative 0; Mismatches 69; Indels 1; Gaps 1;

QY 755 TGTTCACAGAGGCTTCTGACATACAGTCTTTGGGCACTAAGCACTATGCTGTAT 814
DB 12 TTTTCCAGAGGCTTCTGACATGAGTCTGTGAGGGCAGCAATTCCTCTGAT 71
QY 815 TTTTCTGTGCTGTGCTGTGATGACCCCTTACAGAAAGATAGTGGTGAACGAG--GGCG 873
DB 72 TTTTCTGTGCTGTGCTGTGATGACCCCTTACAGAAAGATAGTGGTGAACGAG--GGCG 131
QY 874 GAGCGGCTGTGACATGCTGTGCTGTCTTATCACTTATCATTAAGGAAAGAAAG 933
DB 132 GAGCAGCCCACTTACATGATGTGCTGTCTTATCACTTATCATTAAGGAAAGAAAG 191
QY 934 TGAATGATTCGATACCTGACACTGTAGACTCAGGGAGAAACAAAGTCCGTGACAGC 993
DB 192 TGAATGATTCGATACCTGACACTGTAGACTCAGGGAGAAACAAAGTCCGTGAGAA 251
QY 994 TCTGAGTGAAGCAGGGCTGTCTCTTCTCTCAGAGACG 1032
DB 252 GCTGTCTGTAAAGCTGACAGAGCCCTCCCAAGTAGAC 290

RESULT 11
US-10-433-793-187
; Sequence 187, Application US/10433793
; Publication No. US20040142334A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/433,793
```

```

; CURRENT FILING DATE: 2003-06-06
; NUMBER OF SEQ ID NOS: 212
; SEQ ID NO 187
; LENGTH: 2846
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-433-793-187

Query Match          11.5%; Score 140.6; DB 7; Length 2846;
Best Local Similarity 62.1%; Pred. No. 1.3e-25;
Matches 295; Conservative 0; Mismatches 164; Indels 16; Gaps 4;

QY 522 ATCTTACCTTATATTAACAATAAAGAAAATAAACCAGATCCGATATAGCTGTAAT 581
DB 1530 ATTTTATTTTATTTATTTGTAATATAAATAATGAAGTTTGTATTTATTTGTAATA 1589
QY 582 TTAATTCCT-----AAAGAACGAAACTTTCACATAGCTTTAAATTAAAGTGATTA 634
DB 1590 TTTTATTTTAAAGAAAATAAGAAATTTTATTTGTAATATAAATAAAGTTATTAAT 1649
QY 635 CTCGATCTCTCTGCAAGCTTACCTACAAACGACAGACAGACAGACAGAGCCACGCTAC 694
DB 1650 TTTAGATATTTTGTAGTTAGTACGGTAATAAATTGTTTATGATAAAGATTAATG 1709
QY 695 TCTCT-----AGGAAATAATTAGGCTGCTGCTCTGACATGCCAGGGGCTTGTGCG 747
DB 1710 TTTTATTTTAAAGAAATTTTAAATTTGAGGTGCTGTTTAAAGAAATGTTTAAAGG 1769
QY 748 TGGCTCG-TCGTCAGAAAGGCTTTCGACATACACAGCTCTTGGGACGTAGCACTAT 806
DB 1770 AGATCGGTTTTTTTGAAGGTTTTGTAGTATGCGTTTTGTTGAGGTAAGTAATTTT 1829
QY 807 GCTCTGATTTTCTGCTGCTGCTGCTAGTGACCCCTACAGAAATAGTGGTGAGCCA 866
DB 1830 GTTTGATTTTGTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTGTTG 1889
QY 867 GG-GGGCGAGCGGCTGCGCTGCACATGCTGCTGCTCTTATTAATTTATTAAGG 925
DB 1890 GGAAGGCGAGTAATTTATTAATTAATGTTGTTGTTTATTTATTTATTAAGGA 1949
QY 926 AAGAAAGTATGATTCGATCTGACACTGACACTGACACTGACAGGAGAAACAAGA 980
DB 1950 AAGAAAGTATGATTCGATCTGACACTGATTTGTAAGATTTGGGAGAGAGAAATA 2004

RESULT 12
US-10-433-793-188/c
; Sequence 188, Application US/10433793
; Publication No. US20040142334A1
; GENERAL INFORMATION:
; APPLICANT: Epigenomics AG
; TITLE OF INVENTION: Diagnose von mit Angiogenese assoziierten Krankheiten
; FILE REFERENCE:
; CURRENT APPLICATION NUMBER: US/10/433,793
; CURRENT FILING DATE: 2003-06-06
; NUMBER OF SEQ ID NOS: 212
; SEQ ID NO 188
; LENGTH: 2846
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-433-793-188

Query Match          11.4%; Score 139.4; DB 7; Length 2846;
Best Local Similarity 56.4%; Pred. No. 2.6e-25;
Matches 457; Conservative 0; Mismatches 316; Indels 38; Gaps 9;

QY 206 CCACAGAGTGAATGAGCCCGAGAGAAACCTGATACAGTGAAGAAAGTATGTTTCTT 265
DB 1651 CGACAAAGCAAACTCCGTCTCAAAAAAAAAAAAAAAAAAATTTATTTCTCTCT 1592
```

```

QY 266 TCCTTCGACATACCTTCACTCAAAATATCTATTTACTTAAACAGACATTAATCAGGCCA 325
DB 1591 TCTACACAAACACTTAATTTCTTAATTTCTCTATTAATATATATAACTTAATCAATACA 1532
QY 326 AACCACTTAAGTTTATTTGTAATAGTATTTTGTGTAAGGACAGACATGTGAGTCTG 385
DB 1531 ATATATTTTAAATTAATATCAATATAAATTTTAAATAAATAATCAAA-ACCT 1474
QY 386 AGAAAACTATGTGGTAACTTGAATTAATA-----TATCAAACTGGGTAA 435
DB 1473 CAAAACTAATTTCTTAACGTAACCTTAATAAATAATCTTCTATCAACTATTCGAAA 1414
QY 436 TAAA-----AAAAATGCAATACCTTAATAAATAAATAAATAAATAAATAAATAA 488
DB 1413 TAAAAATCTTTATTAATCTTAATTAATTAATTAATTAATTAATTAATTAATTAAT 1354
QY 489 CTTCCTCTGAAAAAGACA--TTTAAAGGGCTGATCTTAGCCTTATATTTTACAATA 545
DB 1353 TTTCCCTTAAAAAAAACACAGCTTACAAAAACAAACCTTCTCTTATTTATTAATA 1294
QY 546 AAGAAATTAACCAAGTCCGATATAGCTGAATTTATTCCT-----AAAAACA 598
DB 1293 ACATAAATAAACCAGAAATCTTAACCTATTAATTAATTTATTTTCTTAAAAAACA 1234
QY 599 GAACTTTCATATGCTTTAAATTAAGTATTAAGTATTAAGTATTAAGTATTAAGTATTA 657
DB 1233 AAACTTTCATTAATCTTCAACATTAATAATTAATTAATTAATTAATTAATTAATTA 1174
QY 658 CTACAAAGAGAGACAGACAAACAGAGCCCACTACTCTAGGAAA-----TAATTA 711
DB 1173 CGACAAAAATCAATTTCAACAAAAAATAAATCAACTACTCTCTTAAAAAATACCTTA 1114
QY 712 GGGTGGCTCTGACATGCCAGGGGCTTGTGCTGCTGCTG-TCGCCAAGAGGCTT 770
DB 1113 AAATAATACCTTAATAAATACCAAAATCTTAATAAATGATGATTTTCCAAAAAATTT 1054
QY 771 CTGAGTACACAGTCTTTGGGGAGTAAGCACTATGCTGATTTTCTGTGCTGCTG 830
DB 1053 CTACACATTAATCTTAATTAATAAACAACATTTCTTAATTTTCTTAATTAACCTTA 994
QY 831 CTAGTACCCCTACAGAAATAGTGGTGAACC-AGGGGCGAGCGGCTGCTGAC 889
DB 993 CTAAATACCCCTACAAAAAATAACGACTAAACCAAAAAACAAACCAACCTACAC 934
QY 890 ATGCTGCTGCTCTTATCAACTTATCATATATAGGAAAGAAAGTATGATTCGGAATAC 949
DB 933 ATATCTAATCTACTTATCAACTTATCATATATATATATATATATATATATATATAT 874
QY 950 TGACACTGTACACTCAGGGGAGAAACAAGA 980
DB 873 TTAACATAATAAATCTAATAAAAAAAAAAAAAACA 843

RESULT 13
US-10-311-455-866/c
; Sequence 866, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENBROCK, Christian
; TITLE OF INVENTION: Diagnosis of diseases associated with the Immune System by Detect
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
```

```

; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 866
; LENGTH: 10710
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-866

Query Match
Best Local Similarity 4.5%; Score 55; DB 6; Length 10710;
Matches 148; Conservative 0; Mismatches 155; Indels 0; Gaps 0;

QY 256 TATGTTGTTCTCTCGACATATCTTCACTCAATATCTATTGTACTTAACGACAAAT 315
DB 5616 TATTAATCTTAATCTCAAGATATCTTATTAATAAATAAATTCATAAACAACCT 5557

QY 316 AATCAGGCCAAACCACTTAAGTTTATTTAGTATTTTGTAGGACAGACAT 375
DB 5556 AATATATTATTTCTTTTTCATAATACATTTATGATTTTATATATATCTTACATATTT 5497

QY 376 GTGAGTGTGAGAAAACGATGTTGGTAACTTGATTTAATATCAAACTGGGTTAAA 435
DB 5496 TTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5437

QY 436 TAAAAAATGTCATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 495
DB 5436 AATATCTTTCTAATCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5377

QY 496 TTGGAAGAAGACATTTACAAAGGCTGATCTTACCTTTATTTTACATTAAGAAATAA 555
DB 5376 TAAATTTACAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 5317

QY 556 ACC 558
DB 5316 AAC 5314

RESULT 14
US-10-311-455-2114/c
; Sequence 2114, Application US/10311455
; Publication No. US20030143606A1
; GENERAL INFORMATION:
; APPLICANT: OLEK, Alexander
; APPLICANT: PIEPENROCK, Christian
; TITLE OF INVENTION: Diagnosis of Diseases Associated with the Immune System by Determin
; FILE REFERENCE: 5013.1014
; CURRENT APPLICATION NUMBER: US/10/311,455
; CURRENT FILING DATE: 2002-12-16
; PRIOR APPLICATION NUMBER: PCT/EP01/07537
; PRIOR FILING DATE: 2001-07-02
; PRIOR APPLICATION NUMBER: DE 10032529.7
; PRIOR FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: DE 10043826.1
; PRIOR FILING DATE: 2000-09-01
; NUMBER OF SEQ ID NOS: 2424
; SEQ ID NO 2114
; LENGTH: 15698
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-311-455-2114

Query Match
Best Local Similarity 4.5%; Score 54.4; DB 6; Length 15698;
Matches 127; Conservative 0; Mismatches 121; Indels 0; Gaps 0;

QY 416 AATATCAAACTGGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 475
DB 7371 AATTAATAAATAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7312
```

```

QY 476 CAACAAGCTTACTATCCCTTGAAAGACATTTACAAAGGCTGATCTTACCTTTAT 535
DB 7311 AACCAATACTACTTACTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 7252

QY 536 ATTTACAAATAAGAAATTAACCAAGTCCCGATATAGCTATTTTATTTTCTTAAGA 595
DB 7251 AATATCACTTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 7192

QY 596 ACAGAACTTTCACTATGCTTTAAATAATTAAGTATTAATCTAGATCTGCAAGCTTA 655
DB 7191 ACTCTAATCAACAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 7132

QY 656 GCCTACAA 663
DB 7131 CACTACAA 7124

RESULT 15
US-10-602-494-272/c
; Sequence 272, Application US/10602494
; Publication No. US2004026583A1
; GENERAL INFORMATION:
; APPLICANT: Cathy Lofton-Day
; APPLICANT: Andrew Sledziewski
; APPLICANT: Jeff Thomas
; APPLICANT: Robert W. Day
; APPLICANT: Lori Tonnes-Priddy
; APPLICANT: Karen Cardon
; TITLE OF INVENTION: Methods and nucleic acids for the analysis of colorectal cell
; FILE REFERENCE: 47675-45
; CURRENT APPLICATION NUMBER: US/10/602,494
; CURRENT FILING DATE: 2003-06-23
; NUMBER OF SEQ ID NOS: 385
; SEQ ID NO 272
; LENGTH: 2265
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: chemically treated genomic DNA (Homo sapiens)
US-10-602-494-272

Query Match
Best Local Similarity 4.3%; Score 52.6; DB 8; Length 2265;
Matches 115; Conservative 0; Mismatches 104; Indels 0; Gaps 0;

QY 422 AAACGGGTTAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 481
DB 222 AAACCAACTAACTAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 163

QY 482 GACTTACTCTCCCTTGAAAGACATTTACAAAGGCTGATCTTACCTTATATTTAT 541
DB 162 ACATTTACTTAACAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 103

QY 542 AATAAAGAAATTAACCAAGTCCGATATAGCTGTAATTTATTTCTTAAGAACAGAA 601
DB 102 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 43

QY 602 ACTTTCATGCTTTAAATAATTAAGTGTATTAATTAATTAATTAATTAATTAATTA 640
DB 42 ACTAACCAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 4
```

Search completed: December 12, 2005, 15:40:53  
Job time : 1003 secs

**THIS PAGE BLANK (USPTO)**